

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF RECEIPT OF
RECORD COPY

(PCT Rule 24.2(a))

To:

KOIKE, Akira
No.11 Mori Bldg., 6-4, Toranomom 2-
chome
Minato-ku, Tokyo 105-0001
JAPON

Date of mailing (day/month/year) 18 February 2000 (18.02.00)	IMPORTANT NOTIFICATION
Applicant's or agent's file reference SK00PCT13	International application No. PCT/JP00/00423

The applicant is hereby notified that the International Bureau has received the record copy of the international application as detailed below.

Name(s) of the applicant(s) and State(s) for which they are applicants:

SONY CORPORATION (for all designated States except US)
WALKER, Toby et al (for US)

International filing date	:	27 January 2000 (27.01.00)
Priority date(s) claimed	:	29 January 1999 (29.01.99)
Date of receipt of the record copy by the International Bureau	:	14 February 2000 (14.02.00)
List of designated Offices	:	

EP : AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE
National : US


ATTENTION

The applicant should carefully check the data appearing in this Notification. In case of any discrepancy between these data and the indications in the international application, the applicant should immediately inform the International Bureau.

In addition, the applicant's attention is drawn to the information contained in the Annex, relating to:

- ☒ time limits for entry into the national phase
☒ confirmation of precautionary designations
☐ requirements regarding priority documents

A copy of this Notification is being sent to the receiving Office and to the International Searching Authority.

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer:  Shinji IGARASHI
Facsimile No. (41-22) 740.14.35	Telephone No. (41-22) 338.83.38

THIS PAGE BLANK (USPTO)

INFORMATION ON TIME LIMITS FOR ENTERING THE NATIONAL PHASE

The applicant is reminded that the "national phase" must be entered before each of the designated Offices indicated in the Notification of Receipt of Record Copy (Form PCT/IB/301) by paying national fees and furnishing translations, as prescribed by the applicable national laws.

The time limit for performing these procedural acts is **20 MONTHS** from the priority date or, for those designated States which the applicant elects in a demand for international preliminary examination or in a later election, **30 MONTHS** from the priority date, provided that the election is made before the expiration of 19 months from the priority date. Some designated (or elected) Offices have fixed time limits which expire even later than 20 or 30 months from the priority date. In other Offices an extension of time or grace period, in some cases upon payment of an additional fee, is available.

In addition to these procedural acts, the applicant may also have to comply with other special requirements applicable in certain Offices. It is the applicant's responsibility to ensure that the necessary steps to enter the national phase are taken in a timely fashion. Most designated Offices do not issue reminders to applicants in connection with the entry into the national phase.

For detailed information about the procedural acts to be performed to enter the national phase before each designated Office, the applicable time limits and possible extensions of time or grace periods, and any other requirements, see the relevant Chapters of Volume II of the PCT Applicant's Guide. Information about the requirements for filing a demand for international preliminary examination is set out in Chapter IX of Volume I of the PCT Applicant's Guide.

GR and ES became bound by PCT Chapter II on 7 September 1996 and 6 September 1997, respectively, and may, therefore, be elected in a demand or a later election filed on or after 7 September 1996 and 6 September 1997, respectively, regardless of the filing date of the international application. (See second paragraph above.)

Note that only an applicant who is a national or resident of a PCT Contracting State which is bound by Chapter II has the right to file a demand for international preliminary examination.

CONFIRMATION OF PRECAUTIONARY DESIGNATIONS

This notification lists only specific designations made under Rule 4.9(a) in the request. It is important to check that these designations are correct. Errors in designations can be corrected where precautionary designations have been made under Rule 4.9(b). The applicant is hereby reminded that any precautionary designations may be confirmed according to Rule 4.9(c) before the expiration of 15 months from the priority date. If it is not confirmed, it will automatically be regarded as withdrawn by the applicant. There will be no reminder and no invitation. Confirmation of a designation consists of the filing of a notice specifying the designated State concerned (with an indication of the kind of protection or treatment desired) and the payment of the designation and confirmation fees. Confirmation must reach the receiving Office within the 15-month time limit.

REQUIREMENTS REGARDING PRIORITY DOCUMENTS

For applicants who have not yet complied with the requirements regarding priority documents, the following is recalled.

Where the priority of an earlier national, regional or international application is claimed, the applicant must submit a copy of the said earlier application, certified by the authority with which it was filed ("the priority document") to the receiving Office (which will transmit it to the International Bureau) or directly to the International Bureau, before the expiration of 16 months from the priority date, provided that any such priority document may still be submitted to the International Bureau before that date of international publication of the international application, in which case that document will be considered to have been received by the International Bureau on the last day of the 16-month time limit (Rule 17.1(a)).

Where the priority document is issued by the receiving Office, the applicant may, instead of submitting the priority document, request the receiving Office to prepare and transmit the priority document to the International Bureau. Such request must be made before the expiration of the 16-month time limit and may be subjected by the receiving Office to the payment of a fee (Rule 17.1(b)).


If the priority document concerned is not submitted to the International Bureau or if the request to the receiving Office to prepare and transmit the priority document has not been made (and the corresponding fee, if any, paid) within the applicable time limit indicated under the preceding paragraphs, any designated State may disregard the priority claim, provided that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity to furnish the priority document within a time limit which is reasonable under the circumstances.

Where several priorities are claimed, the priority date to be considered for the purposes of computing the 16-month time limit is the filing date of the earliest application whose priority is claimed.

THIS PAGE BLANK (USPTO)

特許協力条約に基づく国際出願願書

副本 - 印刷日時 2000年01月27日 (27.01.2000) 木曜日 15時12分52秒

0	受理官庁記入欄		
0-1	国際出願番号.		
0-2	国際出願日		
0-3	(受付印)		
0-4	様式-PCT/R0/101 この特許協力条約に基づく 国際出願願書は、 右記によって作成された。	PCT-EASY Version 2.90 (updated 15.12.1999)	
0-4-1			
0-5	申立て 出願人は、この国際出願が特許 協力条約に従って処理されるこ とを請求する。		
0-6	出願人によって指定された 受理官庁	日本国特許庁 (R0/JP)	
0-7	出願人又は代理人の書類記 号	SK00PCT13	
I	発明の名称	信号処理方法及び映像音声処理装置	
II	出願人	出願人である (applicant only)	
II-1	この欄に記載した者は	米国を除くすべての指定国 (all designated States except US)	
II-2	右の指定国についての出願人で ある。		
II-4ja	名称	ソニー株式会社	
II-4en	Name	SONY CORPORATION	
II-5ja	あて名:	141-0001 日本国 東京都 品川区 北品川6丁目7番35号	
II-5en	Address:	7-35, Kitashinagawa 6-chome Shinagawa-ku, Tokyo 141-0001 Japan	
II-6	国籍 (国名)	日本国 JP	
II-7	住所 (国名)	日本国 JP	
III-1	その他の出願人又は発明者	出願人及び発明者である (applicant and inventor)	
III-1-1	この欄に記載した者は	米国のみ (US only)	
III-1-2	右の指定国についての出願人で ある。		
III-1-4ja	氏名 (姓名)	ウォーカー トビー	
III-1-4en	Name (LAST, First)	WALKER, Toby	
III-1-5ja	あて名:	141-0001 日本国 東京都 品川区 北品川6丁目7番35号	
III-1-5en	Address:	ソニー株式会社内 c/o SONY CORPORATION 7-35, Kitashinagawa 6-chome Shinagawa-ku, Tokyo 141-0001 Japan	
III-1-6	国籍 (国名)	カナダ CA	
III-1-7	住所 (国名)	日本国 JP	

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

特許協力条約に基づく国際出願願書

副本 - 印刷日時 2000年01月27日 (27.01.2000) 木曜日 15時12分52秒

III-2 III-2-1	その他の出願人又は発明者 この欄に記載した者は	出願人及び発明者である (applicant and inventor)
III-2-2	右の指定国についての出願人である。	米国のみ (US only)
III-2-4ja III-2-4en III-2-5ja	氏名(姓名) Name (LAST, First) あて名:	松原 弘 MATSUBARA, Hiroshi 141-0001 日本国 東京都 品川区 北品川6丁目7番35号 ソニー株式会社内
III-2-5en	Address:	c/o SONY CORPORATION 7-35, Kitashinagawa 6-chome Shinagawa-ku, Tokyo 141-0001 Japan
III-2-6	国籍 (国名)	日本国 JP
III-2-7	住所 (国名)	日本国 JP
IV-1	代理人又は共通の代表者、 通知のあて名 下記の者は国際機関において右記のごとく出願人のために行動する。	代理人 (agent)
IV-1-1ja IV-1-1en IV-1-2ja	氏名(姓名) Name (LAST, First) あて名:	小池 晃 KOIKE, Akira 105-0001 日本国 東京都 港区 虎ノ門二丁目6番4号 第11森ビル
IV-1-2en	Address:	No.11 Mori Bldg., 6-4, Toranomon 2-chome Minato-ku, Tokyo 105-0001 Japan
IV-1-3	電話番号	03-3508-8266
IV-1-4	ファクシミリ番号	03-3508-0439
IV-2	その他の代理人	筆頭代理人と同じあて名を有する代理人 (additional agent(s) with same address as first named agent)
IV-2-1ja IV-2-1en	氏名 Name(s)	田村 栄一; 伊賀 誠司 TAMURA, Eiichi; IGA, Seiji
V	国の指定	
V-1	広域特許 (他の種類の保護又は取扱いを 求める場合には括弧内に記載する。)	EP: AT BE CH&LI CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE 及びヨーロッパ特許条約と特許協力条約の締約国 である他の国
V-2	国内特許 (他の種類の保護又は取扱いを 求める場合には括弧内に記載する。)	US

THIS PAGE BLANK (USPTO)

特許協力条約に基づく国際出願願書

副本 - 印刷日時 2000年01月27日 (27. 01. 2000) 木曜日 15時12分52秒

V-5	指定の確認の宣言 出願人は、上記の指定に加えて、規則4.9(b)の規定に基づき、特許協力条約のもとで認められる他の全ての国の指定を行う。ただし、V-6欄に示した国の指定を除く。出願人は、これらの追加される指定が確認を条件としていること、並びに優先日から15月が経過する前にその確認がなされない指定は、この期間の経過時に、出願人によって取り下げられたものとみなされることを宣言する。	
V-6	指定の確認から除かれる国	なし (NONE)
VI-1	先の国内出願に基づく優先権主張	
VI-1-1	先の出願日	1999年01月29日 (29. 01. 1999)
VI-1-2	先の出願番号	平成 1 1 年特許願第 0 2 3 0 6 4
VI-1-3	国名	日本国 JP
VII-1	特定された国際調査機関 (ISA)	日本国特許庁 (ISA/JP)
VIII	照合欄	用紙の枚数
VIII-1	願書	4
VIII-2	明細書	36
VIII-3	請求の範囲	4
VIII-4	要約	1
VIII-5	図面	6
VIII-7	合計	51
VIII-8	添付書類	添付
VIII-8	手数料計算用紙	✓
VIII-9	別個の記名押印された委任状	✓
VIII-12	優先権証明書	優先権証明書 VI-1
VIII-16	PCT-EASYディスク	-
VIII-17	その他	納付する手数料に相当する特許印紙を貼付した書面
VIII-18	要約書とともに提示する図の番号	3
VIII-19	国際出願の使用言語名:	日本語 (Japanese)
IX	提出者の記名押印	
IX-1	氏名 (姓名)	
IX-2	権限	

受理官庁記入欄

10-1	国際出願として提出された書類の実際の受理の日	
10-2	図面:	
10-2-1	受理された	
10-2-2	不足図面がある	

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

特許協力条約に基づく国際出願願書

副本 - 印刷日時 2000年01月27日 (27.01.2000) 木曜日 15時12分52秒

10-3	国際出願として提出された書類を補完する書類又は図面であつてその後期間内に提出されたものの実際の受理の日 (訂正日)	
10-4	特許協力条約第11条(2)に基づく必要な補完の期間内の受理の日	
10-5	出願人により特定された国際調査機関	ISA/JP
10-6	調査手数料未払いにつき、国際調査機関に調査用写しを送付していない	

国際事務局記入欄

11-1	記録原本の受理の日	
------	-----------	--

THIS PAGE BLANK (USPTO)

E P



P C T

国際調査報告

(法8条、法施行規則第40、41条)
[PCT18条、PCT規則43、44]

出願人又は代理人 の書類記号 SK00PCT13	今後の手続きについては、国際調査報告の送付通知様式(PCT/ISA/220)及び下記5を参照すること。	
国際出願番号 PCT/JP00/00423	国際出願日 (日.月.年) 27.01.00	優先日 (日.月.年) 29.01.99
出願人(氏名又は名称) ソニー株式会社		

国際調査機関が作成したこの国際調査報告を法施行規則第41条(PCT18条)の規定に従い出願人に送付する。
この写しは国際事務局にも送付される。

この国際調査報告は、全部で 3 ページである。

☐ この調査報告に引用された先行技術文献の写しも添付されている。

1. 国際調査報告の基礎

a. 言語は、下記に示す場合を除くほか、この国際出願がされたものに基づき国際調査を行った。

☐ この国際調査機関に提出された国際出願の翻訳文に基づき国際調査を行った。

b. この国際出願は、ヌクレオチド又はアミノ酸配列を含んでおり、次の配列表に基づき国際調査を行った。

☐ この国際出願に含まれる書面による配列表

☐ この国際出願と共に提出されたフレキシブルディスクによる配列表

☐ 出願後に、この国際調査機関に提出された書面による配列表

☐ 出願後に、この国際調査機関に提出されたフレキシブルディスクによる配列表

☐ 出願後に提出した書面による配列表が出願時における国際出願の開示の範囲を超える事項を含まない旨の陳述書の提出があった。

☐ 書面による配列表に記載した配列とフレキシブルディスクによる配列表に記録した配列が同一である旨の陳述書の提出があった。

2. ☐ 請求の範囲の一部の調査ができない(第I欄参照)。

3. ☐ 発明の単一性が欠如している(第II欄参照)。

4. 発明の名称は ☒ 出願人が提出したものを承認する。

☐ 次に示すように国際調査機関が作成した。

5. 要約は ☒ 出願人が提出したものを承認する。

☐ 第III欄に示されているように、法施行規則第47条(PCT規則38.2(b))の規定により国際調査機関が作成した。出願人は、この国際調査報告の発送の日から1カ月以内にこの国際調査機関に意見を提出することができる。

6. 要約書とともに公表される図は、

第 3 図とする。 ☒ 出願人が示したとおりである。

☐ なし

☐ 出願人は図を示さなかった。

☐ 本図は発明の特徴を一層よく表している。

THIS PAGE BLANK (USPTO)

A. 発明の属する分野の分類 (国際特許分類 (IPC)) Int. Cl ⁷ H04N7/32 Int. Cl ⁷ H04N5/92 Int. Cl ⁷ G06T7/00		
B. 調査を行った分野 調査を行った最小限資料 (国際特許分類 (IPC)) Int. Cl ⁷ H04N5/91~5/956 Int. Cl ⁷ H04N7/24~7/68 Int. Cl ⁷ H04N5/222~5/257 Int. Cl ⁷ G06T7/00 Int. Cl ⁷ H04N5/262~5/28		
最小限資料以外の資料で調査を行った分野に含まれるもの 日本国実用新案公報 1922-1996年 日本国公開実用新案公報 1971-2000年 日本国登録実用新案公報 1994-2000年 日本国実用新案登録公報 1996-2000年		
国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)		
C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	JP, 10-257436, A (松下温, 岡田謙一) 25. 9月. 1998 (25. 09. 98) 全文, 第1-20図 (ファミリーなし)	1-13
A	JP, 7-193748, A (日本電信電話株式会社) 28. 7月. 1995 (28. 07. 95) 全文, 第1-14図 (ファミリーなし)	1-13
<input checked="" type="checkbox"/> C欄の続きにも文献が列挙されている。 <input type="checkbox"/> パテントファミリーに関する別紙を参照。		
* 引用文献のカテゴリー 「A」 特に関連のある文献ではなく、一般的技術水準を示すもの 「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの 「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す) 「O」 口頭による開示、使用、展示等に言及する文献 「P」 国際出願日前で、かつ優先権の主張の基礎となる出願日の後に公表された文献 「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの 「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの 「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの 「&」 同一パテントファミリー文献		
国際調査を完了した日 18. 04. 00	国際調査報告の発送日 02.05.00	
国際調査機関の名称及びあて先 日本国特許庁 (ISA/J P) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官 (権限のある職員) 安田 太 電話番号 03-3581-1101 内線 3531	

THIS PAGE BLANK (USPTO)

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	EP, 711078, A (松下電器産業株式会社) 8. 5月. 1996 (08. 05. 96) 全文, 全図 & JP, 8-181995, A & US, 5751377, A	1-13

THIS PAGE BLANK (USPTO)

PCT COOPERATION TREATY

PCT

NOTIFICATION CONCERNING
SUBMISSION OR TRANSMITTAL
OF PRIORITY DOCUMENT

(PCT Administrative Instructions, Section 411)

From the INTERNATIONAL BUREAU

To:

KOIKE, Akira
No.11 Mori Bldg., 6-4, Toranomom 2-
chome
Minato-ku, Tokyo 105-0001
JAPON

Date of mailing (day/month/year) 18 February 2000 (18.02.00)	IMPORTANT NOTIFICATION
Applicant's or agent's file reference SK00PCT13	
International application No. PCT/JP00/00423	
International publication date (day/month/year) Not yet published	
Applicant SONY CORPORATION et al	International filing date (day/month/year) 27 January 2000 (27.01.00) Priority date (day/month/year) 29 January 1999 (29.01.99)

- The applicant is hereby notified of the date of receipt (except where the letters "NR" appear in the right-hand column) by the International Bureau of the priority document(s) relating to the earlier application(s) indicated below. Unless otherwise indicated by an asterisk appearing next to a date of receipt, or by the letters "NR", in the right-hand column, the priority document concerned was submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b).
- This updates and replaces any previously issued notification concerning submission or transmittal of priority documents.
- An asterisk(*) appearing next to a date of receipt, in the right-hand column, denotes a priority document submitted or transmitted to the International Bureau but not in compliance with Rule 17.1(a) or (b). In such a case, **the attention of the applicant is directed to Rule 17.1(c)** which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.
- The letters "NR" appearing in the right-hand column denote a priority document which was not received by the International Bureau or which the applicant did not request the receiving Office to prepare and transmit to the International Bureau, as provided by Rule 17.1(a) or (b), respectively. In such a case, **the attention of the applicant is directed to Rule 17.1(c)** which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.

<u>Priority date</u>	<u>Priority application No.</u>	<u>Country or regional Office or PCT receiving Office</u>	<u>Date of receipt of priority document</u>
29 Janu 1999 (29.01.99)	11/23064	JP	14 Febr 2000 (14.02.00)

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. (41-22) 740.14.35

Authorized officer

Shinji IGARASHI

Telephone No. (41-22) 338.83.38

THIS PAGE BLANK (USPTO)

PATENT COOPERATION TREATY

PCT

NOTICE INFORMING THE APPLICANT OF THE COMMUNICATION OF THE INTERNATIONAL APPLICATION TO THE DESIGNATED OFFICES

(PCT Rule 47.1(c), first sentence)

From the INTERNATIONAL BUREAU

To:

KOIKE, Akira
11 Mori Bldg., 6-4, Toranomom 2-
chome
Minato-ku, Tokyo 105-0001
JAPON

Date of mailing (day/month/year) 03 August 2000 (03.08.00)		
Applicant's or agent's file reference SK00PCT13		IMPORTANT NOTICE
International application No. PCT/JP00/00423	International filing date (day/month/year) 27 January 2000 (27.01.00)	
Priority date (day/month/year) 29 January 1999 (29.01.99)		
Applicant SONY CORPORATION et al		

1. Notice is hereby given that the International Bureau has communicated, as provided in Article 20, the international application to the following designated Offices on the date indicated above as the date of mailing of this Notice:

US

In accordance with Rule 47.1(c), third sentence, those Offices will accept the present Notice as conclusive evidence that the communication of the international application has duly taken place on the date of mailing indicated above and no copy of the international application is required to be furnished by the applicant to the designated Office(s).

2. The following designated Offices have waived the requirement for such a communication at this time:

EP

The communication will be made to those Offices only upon their request. Furthermore, those Offices do not require the applicant to furnish a copy of the international application (Rule 49.1(a-bis)).

3. Enclosed with this Notice is a copy of the international application as published by the International Bureau on 03 August 2000 (03.08.00) under No. WO 00/45604

REMINDER REGARDING CHAPTER II (Article 31(2)(a) and Rule 54.2)

If the applicant wishes to postpone entry into the national phase until 30 months (or later in some Offices) from the priority date, a demand for international preliminary examination must be filed with the competent International Preliminary Examining Authority before the expiration of 19 months from the priority date.

It is the applicant's sole responsibility to monitor the 19-month time limit.

Note that only an applicant who is a national or resident of a PCT Contracting State which is bound by Chapter II has the right to file a demand for international preliminary examination.

REMINDER REGARDING ENTRY INTO THE NATIONAL PHASE (Article 22 or 39(1))

If the applicant wishes to proceed with the international application in the national phase, he must, within 20 months or 30 months, or later in some Offices, perform the acts referred to therein before each designated or elected Office.

For further important information on the time limits and acts to be performed for entering the national phase, see the Annex to Form PCT/IB/301 (Notification of Receipt of Record Copy) and Volume II of the PCT Applicant's Guide.

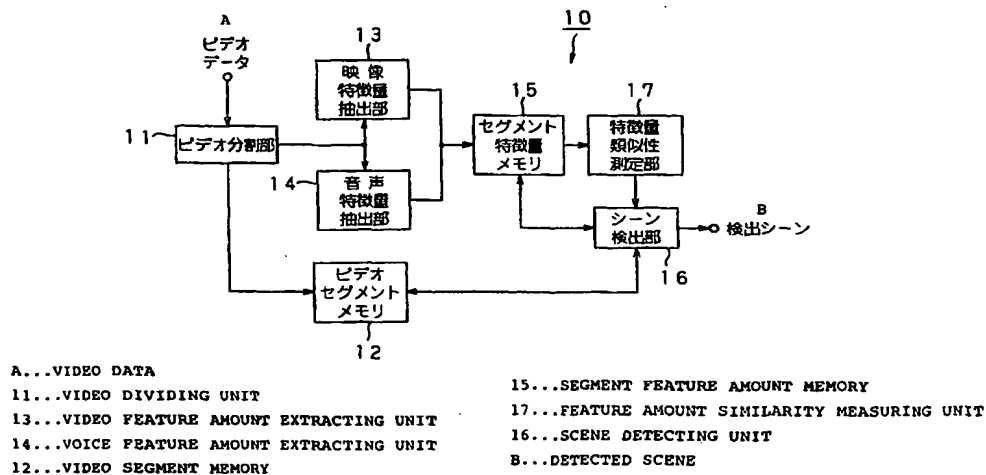
The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer J. Zahra
Facsimile No. (41-22) 740.14.35	Telephone No. (41-22) 338.83.38

THIS PAGE BLANK (USPTO)

(51) 国際特許分類7 H04N 7/32, 5/92, G06T 7/00		A1	(11) 国際公開番号 WO00/45604
			(43) 国際公開日 2000年8月3日(03.08.00)
(21) 国際出願番号 PCT/JP00/00423		(81) 指定国 US, 欧州特許 (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE)	
(22) 国際出願日 2000年1月27日(27.01.00)		添付公開書類 国際調査報告書	
(30) 優先権データ 特願平11/23064 1999年1月29日(29.01.99) JP			
(71) 出願人 (米国を除くすべての指定国について) ソニー株式会社(SONY CORPORATION)[JP/JP] 〒141-0001 東京都品川区北品川6丁目7番35号 Tokyo, (JP)			
(72) 発明者; および			
(75) 発明者/出願人 (米国についてののみ) ウォーカー トビー(WALKER, Toby)[CA/JP] 松原 弘(MATSUBARA, Hiroshi)[JP/JP] 〒141-0001 東京都品川区北品川6丁目7番35号 ソニー株式会社内 Tokyo, (JP)			
(74) 代理人 小池 晃, 外(KOIKE, Akira et al.) 〒105-0001 東京都港区虎ノ門二丁目6番4号 第11森ビル Tokyo, (JP)			

(54)Title: SIGNAL PROCESSING METHOD AND VIDEO/VOICE PROCESSING DEVICE

(54)発明の名称 信号処理方法及び映像音声処理装置



(57) Abstract

A video/voice processing device (10) provided with a scene detection unit (16) which uses feature amounts extracted from video segments and or voice segments divided from an input video data stream and a measuring reference for measuring, using the feature amounts, similarities between pairs of video segments and/or voice segments calculated for each feature amount, so as to detect, out of the video segments and/or voice segments, two video segments and/or voice segments that have a mutual time-wise distance within a preset time threshold and a mutual unsimilarity up to a preset unsimilarity threshold and to integrate the segments into a scene that reflects the semantic structure of the video data contents and consists of time-wise continuous video segments and/or voice segments.

映像音声処理装置 10 は、入力したビデオデータのストリームから分割された映像セグメント及び／又は音声セグメントから抽出された特徴量と、この特徴量を用いて、各特徴量毎に計算された、映像セグメント及び／又は音声セグメントの対の間の類似性を測定する測定基準とを用いて、映像セグメント及び／又は音声セグメントのうち、互いの時間的距離が所定の時間閾値以内であるとともに、互いの非類似性が所定の非類似性閾値以下である 2 つの映像セグメント及び／又は音声セグメントを検出し、ビデオデータの内容の意味構造を反映し時間的に連続する映像セグメント及び／又は音声セグメントからなるシーンにまとめるシーン検出部 16 を備える。

PCTに基づいて公開される国際出願のパンフレット第一頁に掲載されたPCT加盟国を同定するために使用されるコード(参考情報)

AE	アラブ首長国連邦	DM	ドミニカ	KZ	カザフスタン	RU	ロシア
AG	アンティグア・バーブーダ	DZ	アルジェリア	LC	セントルシア	SD	スーダン
AL	アルバニア	EE	エストニア	LI	リヒテンシュタイン	SE	スウェーデン
AM	アルメニア	ES	スペイン	LK	スリ・ランカ	SG	シンガポール
AT	オーストリア	FI	フィンランド	LR	リベリア	SI	スロヴェニア
AU	オーストラリア	FR	フランス	LS	レソト	SK	スロヴァキア
AZ	アゼルバイジャン	GA	ガボン	LT	リトアニア	SL	シエラ・レオネ
BA	ボスニア・ヘルツェゴビナ	GB	英国	LU	ルクセンブルグ	SN	セネガル
BB	バルバドス	GD	グレナダ	LV	ラトヴィア	SZ	スワジランド
BE	ベルギー	GE	グルジア	MA	モロッコ	TD	チャード
BF	ブルキナ・ファソ	GH	ガーナ	MC	モナコ	TG	トーゴ
BG	ブルガリア	GM	ガンビア	MD	モルドヴァ	TJ	タジキスタン
BJ	ベナン	GN	ギニア	MG	マダガスカル	TM	トルクメニスタン
BR	ブラジル	GR	ギリシャ	MK	マケドニア旧ユーゴスラヴィア	TR	トルコ
BY	ベラルーシ	CW	キニア・ビスオ		共和国	TT	トリニダード・トバゴ
CA	カナダ	HR	クロアチア	ML	マリ	TT	トリニダード・トバゴ
CF	中央アフリカ	HU	ハンガリー	MN	モンゴル	TZ	タンザニア
CG	コンゴ	ID	インドネシア	MR	モーリタニア	UA	ウクライナ
CH	スイス	IE	アイルランド	MW	マラウイ	UG	ウガンダ
CI	コートジボアール	IL	イスラエル	MX	メキシコ	US	米国
CN	中国	IN	インド	MZ	モザンビーク	UZ	ウズベキスタン
CM	カメルーン	IS	アイスランド	NE	ニジェール	VN	ヴェトナム
CR	コスタ・リカ	IT	イタリア	NL	オランダ	YU	ユーゴスラヴィア
CU	キューバ	JP	日本	NO	ノールウェー	ZA	南アフリカ共和国
CY	キプロス	KE	ケニア	NZ	ニュージーランド	ZW	ジンバブエ
CZ	チェッコ	KG	キルギスタン	PL	ポーランド		
DE	ドイツ	KP	北朝鮮	PT	ポルトガル		
DK	デンマーク	KR	韓国	RO	ルーマニア		

明細書

信号処理方法及び映像音声処理装置

技術分野

本発明は、信号の基礎となる意味構造を反映するパターンを検出して解析する信号処理方法及びビデオ信号の基礎となる意味構造を反映する映像及び／又は音声のパターンを検出して解析する映像音声処理装置に関する。

背景技術

例えばビデオデータに録画されたテレビ番組といった大量の異なる映像データにより構成される映像アプリケーションの中から、興味のある部分等の所望の部分を探して再生したい場合がある。

このように、所望の映像内容を抽出するための一般的な技術としては、アプリケーションの主要場面を描いた一連の映像を並べて作成されたパネルであるストーリーボードがある。このストーリーボードは、ビデオデータをいわゆるショットに分解し、各ショットにおいて代表される映像を表示したものである。このような映像抽出技術は、そのほとんどが、例えば “G. Ahanger and T.D.C. Little, A survey of technologies for parsing and indexing digital video, J. of Visual Communication and Image Representation 7: 28-4, 1996” に記載されているように、ビデオデータからショット

を自動的に検出して抽出するものである。

ところで、例えば代表的な30分のテレビ番組中には、数百ものショットが含まれている。そのため、上述した従来の映像抽出技術においては、ユーザが抽出された膨大な数のショットを並べたストーリーボードを調べる必要があり、このようなストーリーボードを理解するにはユーザに大きな負担を強いる必要があった。また、従来の映像抽出技術においては、例えば話し手の変化に応じて交互に2者を撮影した会話場面におけるショットは、冗長のものが多いという問題があった。このように、ショットは、ビデオ構造を抽出する対象としては階層が低すぎて無駄な情報量が多く、このようなショットを抽出する従来の映像抽出技術は、ユーザにとって利便のよいものとはいえなかった。

また、他の映像抽出技術としては、例えば“A. Merlino, D. Morrey and M. Maybury, Broadcast news navigation using story segmentation, Proc. of ACM Multimedia 97, 1997”や特開平10-136297号公報に記載されているように、ニュースやフットボールゲームといった特定の内容ジャンルに関する非常に専門的な知識を用いるものがある。しかしながら、この従来の映像抽出技術は、目的のジャンルに関しては良好な結果を得ることができるものの他のジャンルには全く役に立たず、さらにジャンルに限定される結果、容易に一般化することができないという問題があった。

さらに、他の映像抽出技術としては、例えばU.S. Patent #5,708,767号公報に記載されているように、いわゆるストーリーユニットを抽出するものがある。しかしながら、この従来の映像抽出技術は、完全に自動化されたものではなく、どのショットが同じ内容を示す

ものであるかを決定するために、ユーザの介入が必要であった。また、この従来の映像抽出技術は、処理に要する計算が複雑であるとともに、適用対象として映像情報のみに限定されるといった問題もあった。

さらにまた、他の映像抽出技術としては、例えば特開平 9 - 2 1 4 8 7 9 号公報に記載されているように、ショット検出と無音部分検出とを組み合わせることによりショットを識別するものがある。しかしながら、この従来の映像抽出技術は、無音部分がショット境界に対応した場合のみに限定されたものであった。

また、他の映像抽出技術としては、例えば “H. Aoki, S. Shimotsuji and O. Hori, A shot classification method to select effective key-frames for video browsing, IPSJ Human Interface SIG Notes, 7:43-50, 1996” や特開平 9 - 9 3 5 8 8 号公報に記載されているように、ストーリーボードにおける表示の冗長を低減するために、反復された類似ショットを検出するものがある。しかしながら、この従来の映像抽出技術は、映像情報のみに適用できるものであり、音声情報に適用できるものではなかった。

発明の開示

本発明は、このような実情に鑑みてなされたものであり、上述した従来の映像抽出技術の問題を解決し、種々のビデオデータにおける高いレベルのビデオ構造を抽出する信号処理方法及び映像音声処理装置を提供することを目的とするものである。

上述した目的を達成する本発明にかかる信号処理方法は、供給さ

れた信号の内容の意味構造を反映するパターンを検出して解析する信号処理方法であって、信号を構成する連続したフレームのひと続きから形成されるセグメントから、その特徴を表す少なくとも1つ以上の特徴量を抽出する特徴量抽出工程と、特徴量を用いて、特徴量のそれぞれ毎に、セグメントの対の間の類似性を測定する測定基準を算出して、この測定基準によりセグメントの対の間の類似性を測定する類似性測定工程と、特徴量と測定基準とを用いて、セグメントのうち、互いの時間的距離が所定の時間閾値以内であるとともに、互いの非類似性が所定の非類似性閾値以下である2つのセグメントを検出し、信号の内容の意味構造を反映し時間的に連続するセグメントからなるシーンにまとめるグループ化工程とを備えることを特徴としている。

このような本発明にかかる信号処理方法は、信号において類似したセグメントを検出してシーンにまとめる。

また、上述した目的を達成する本発明にかかる映像音声処理装置は、供給されたビデオ信号の内容の意味構造を反映する映像及び／又は音声のパターンを検出して解析する映像音声処理装置であって、ビデオ信号を構成する連続した映像及び／又は音声フレームのひと続きから形成される映像及び／又は音声セグメントから、その特徴を表す少なくとも1つ以上の特徴量を抽出する特徴量抽出手段と、特徴量を用いて、特徴量のそれぞれ毎に、映像及び／又は音声セグメントの対の間の類似性を測定する測定基準を算出して、この測定基準により映像及び／又は音声セグメントの対の間の類似性を測定する類似性測定手段と、特徴量と測定基準とを用いて、映像及び／又は音声セグメントのうち、互いの時間的距離が所定の時間閾値以

内であるとともに、互いの非類似性が所定の非類似性閾値以下である2つの映像及び／又は音声セグメントを検出し、ビデオ信号の内容の意味構造を反映し時間的に連続する映像及び／又は音声セグメントからなるシーンにまとめるグループ化手段とを備えることを特徴としている。

このような本発明にかかる映像音声処理装置は、ビデオ信号において類似した映像及び／又は音声セグメントを検出してまとめ、シーンとして出力する。

図面の簡単な説明

図1は、本発明において適用するビデオデータの構成を説明する図であって、モデル化したビデオデータの構造を説明する図である。

図2は、シーンを説明する図である。

図3は、本発明の実施の形態として示す映像音声処理装置の構成を説明するブロック図である。

図4は、同映像音声処理装置において、シーンを検出してグループ化する際の一連の工程を説明するフローチャートである。

図5は、同映像音声処理装置における動的特徴量サンプリング処理を説明する図である。

図6は、非類似性閾値を説明する図である。

図7は、時間閾値を説明する図である。

図8は、同映像音声処理装置において、セグメントをグループ化する際の一連の工程を説明するフローチャートである。

発明を実施するための最良の形態

以下、本発明を適用した具体的な実施の形態について図面を参照しながら詳細に説明する。

本発明を適用した実施の形態は、録画されたビデオデータから所望の内容を自動的に探し出して抽出する映像音声処理装置である。この映像音声処理装置の具体的な説明を行う前に、ここではまず本発明において対象とするビデオデータに関する説明を行う。

本発明において対象とするビデオデータについては、図1に示すようにモデル化し、フレーム、セグメント、シーンの3つのレベルに階層化された構造を有するものとする。すなわち、ビデオデータは、最下位層において、一連のフレームにより構成される。また、ビデオデータは、フレームの1つ上の階層として、連続するフレームのひと続きから形成されるセグメントにより構成される。さらに、ビデオデータは、最上位層において、このセグメントを意味のある関連に基づきまとめて形成されるシーンにより構成される。

このビデオデータは、映像及び音声の両方の情報を含む。すなわち、このビデオデータにおけるフレームには、単一の静止画像である映像フレームと、一般に数十～数百ミリ秒／長といった短時間においてサンプルされた音声情報を表す音声フレームとが含まれる。

また、セグメントは、単一のカメラにより連続的に撮影された映像フレームのひと続きから構成され、一般にはショットと呼ばれる。そして、このセグメントには、映像セグメント及び／又は音声セグメントが含まれ、ビデオ構造における基本単位となる。これらのセ

グメントの中で、特に音声セグメントについては、多くの定義が可能であり、例として次に示すようなものが考えられる。まず、音声セグメントは、一般によく知られている方法により検出されたビデオデータ中の無音期間により境界を定められて形成される場合がある。また、音声セグメントは、“D. Kimber and L. Wilcox, Acoustic Segmentation for Audio Browsers, Xerox Parc Technical Report”に記載されているように、例えば、音声、音楽、ノイズ、無音等のように少数のカテゴリに分類された音声フレームのひと続きから形成される場合もある。さらに、音声セグメントは、“S. Pfeiffer, S. Fischer and E. Wolfgang, Automatic Audio Content Analysis, Proceeding of ACM Multimedia 96, Nov. 1996, pp21-30”に記載されているように、2枚の連続する音声フレーム間の或る特徴における大きな変化を音声カット点として検出し、この音声カット点に基づいて決定される場合もある。

さらに、シーンは、ビデオデータの内容を意味に基づくより高いレベルで記述するために、映像セグメント（ショット）検出或いは音声セグメント検出により得られたセグメントを、例えばセグメント内の知覚的アクティビティ量といったセグメントの特徴を表す特徴量を用いて意味のあるまとまりにグループ化したものである。シーンは、主観的なものであり、ビデオデータの内容或いはジャンルに依存するが、ここでは、その特徴量が互いに類似性を示す映像セグメント又は音声セグメントの反復パターンをグループ化したものとする。具体的には、図2に示すように、2人の話し手が互いに会話をしている場面で、映像セグメントは、話し手に応じて交互に現れる。このような反復パターンを有するビデオデータにおいて、一方

の話し手における一連の映像セグメントAと、他方の話し手における一連の映像セグメントBとは、それぞれ、グループにまとめられて1つのシーンを構成する。このような反復パターンは、ビデオデータにおける高いレベルでの意味のある構造と非常に関係があり、シーンは、このようなビデオデータにおける高いレベルでの意味を持ったまとまりを示すものである。

本発明を適用した実施の形態として図3に示す映像音声処理装置10は、上述したビデオデータにおけるセグメントの特徴量を用いてセグメント間の類似性を測定し、これらのセグメントをシーンにまとめてビデオ構造を自動的に抽出するものであり、映像セグメント及び音声セグメントの両方に適用できるものである。

映像音声処理装置10は、同図に示すように、入力したビデオデータのストリームを映像、音声又はこれらの両方のセグメントに分割するビデオ分割部11と、ビデオデータの分割情報を記憶するビデオセグメントメモリ12と、各映像セグメントにおける特徴量を抽出する特徴量抽出手段である映像特徴量抽出部13と、各音声セグメントにおける特徴量を抽出する特徴量抽出手段である音声特徴量抽出部14と、映像セグメント及び音声セグメントの特徴量を記憶するセグメント特徴量メモリ15と、映像セグメント及び音声セグメントをシーンにまとめるグループ化手段であるシーン検出部16と、2つのセグメント間の類似性を測定する類似性測定手段である特徴量類似性測定部17とを備える。

ビデオ分割部11は、例えば、MPEG1 (Moving Picture Experts Group phase 1) やMPEG2 (Moving Picture Experts Group phase 2)、或いはいわゆるDV (Digital Video) のような圧縮

ビデオデータフォーマットを含む種々のデジタル化されたフォーマットにおける映像データと音声データとからなるビデオデータのストリームを入力し、このビデオデータを映像、音声又はこれらの両方のセグメントに分割するものである。このビデオ分割部 11 は、入力したビデオデータが圧縮フォーマットであった場合、この圧縮ビデオデータを完全伸張することなく直接処理することができる。ビデオ分割部 11 は、入力したビデオデータを処理し、映像セグメントと音声セグメントとに分割する。また、ビデオ分割部 11 は、入力したビデオデータを分割した結果である分割情報を後段のビデオセグメントメモリ 12 に供給する。さらに、ビデオ分割部 11 は、映像セグメントと音声セグメントとに応じて、分割情報を後段の映像特徴量抽出部 13 及び音声特徴量抽出部 14 に供給する。

ビデオセグメントメモリ 12 は、ビデオ分割部 11 から供給されたビデオデータの分割情報を記憶する。また、ビデオセグメントメモリ 12 は、後述するシーン検出部 16 からの問い合わせに応じて、分割情報をシーン検出部 16 に供給する。

映像特徴量抽出部 13 は、ビデオ分割部 11 によりビデオデータを分割して得た各映像セグメント毎の特徴量を抽出する。映像特徴量抽出部 13 は、圧縮映像データを完全伸張することなく直接処理することができる。映像特徴量抽出部 13 は、抽出した各映像セグメントの特徴量を後段のセグメント特徴量メモリ 15 に供給する。

音声特徴量抽出部 14 は、ビデオ分割部 11 によりビデオデータを分割して得た各音声セグメント毎の特徴量を抽出する。音声特徴量抽出部 14 は、圧縮音声データを完全伸張することなく直接処理することができる。音声特徴量抽出部 14 は、抽出した各音声セグ

メントの特徴量を後段のセグメント特徴量メモリ 15 に供給する。

セグメント特徴量メモリ 15 は、映像特徴量抽出部 13 及び音声特徴量抽出部 14 からそれぞれ供給された映像セグメント及び音声セグメントの特徴量を記憶する。セグメント特徴量メモリ 15 は、後述する特徴量類似性測定部 17 からの問い合わせに応じて、記憶している特徴量やセグメントを特徴量類似性測定部 17 に供給する。

シーン検出部 16 は、ビデオセグメントメモリ 12 に保持された分割情報と、1 対のセグメント間の類似性とを用いて、映像セグメント及び音声セグメントをそれぞれシーンにまとめる。シーン検出部 16 は、グループ内の各セグメントから開始して、セグメント群の中から類似しているセグメントの反復パターンを検出し、このようなセグメントを同一シーンとしてまとめてグループ化する。このシーン検出部 16 は、或るシーンにおけるセグメントをまとめてグループを徐々に大きくしていき、全てのセグメントをグループ化するまで処理を行い、最終的に検出シーンを生成して出力する。シーン検出部 16 は、特徴量類似性測定部 17 を用いて、2 つのセグメントがどの程度類似しているかを判断する。

特徴量類似性測定部 17 は、2 つのセグメント間の類似性を測定する。特徴量類似性測定部 17 は、或るセグメントに関する特徴量を検索するようにセグメント特徴量メモリ 15 に問いかける。

時間的に近接して反復している類似したセグメントは、ほぼ同一シーンの一部であるため、映像音声処理装置 10 は、このようなセグメントを検出してグループ化していくことによって、シーンを検出する。このような映像音声処理装置 10 は、図 4 に概略を示すような一連の処理を行うことによって、シーンを検出する。

まず、映像音声処理装置 10 は、同図に示すように、ステップ S 1 において、ビデオ分割を行う。すなわち、映像音声処理装置 10 は、ビデオ分割部 11 に入力されたビデオデータを映像セグメント又は音声セグメントのいずれか、或いは可能であればその両方に分割する。映像音声処理装置 10 は、適用するビデオ分割方法に特に前提要件を設けない。例えば、映像音声処理装置 10 は、“G. Ahanger and T.D.C. Little, A survey of technologies for parsing and indexing digital video, J. of Visual Communication and Image Representation 7:28-4, 1996” に記載されているような方法によりビデオ分割を行う。このようなビデオ分割の方法は、当該技術分野ではよく知られたものであり、映像音声処理装置 10 は、いかなるビデオ分割方法も適用できるものとする。

続いて、映像音声処理装置 10 は、ステップ S 2 において、特徴量の抽出を行う。すなわち、映像音声処理装置 10 は、映像特徴量抽出部 13 や音声特徴量抽出部 14 によって、そのセグメントの特徴を表す特徴量を計算する。映像音声処理装置 10 においては、例えば、各セグメントの時間長、カラーヒストグラムやテクスチャフィーチャといった映像特徴量や、周波数解析結果、レベル、ピッチといった音声特徴量や、アクティビティ測定結果等が、適用可能な特徴量として計算される。勿論、映像音声処理装置 10 は、適用可能な特徴量としてこれらに限定されるものではない。

続いて、映像音声処理装置 10 は、ステップ S 3 において、特徴量を用いたセグメントの類似性測定を行う。すなわち、映像音声処理装置 10 は、特徴量類似性測定部 17 により非類似性測定を行い、その測定基準によって、2つのセグメントがどの程度類似している

かを測定する。映像音声処理装置 10 は、先のステップ S 2 において抽出した特徴量を用いて、非類似性測定基準を計算する。

そして、映像音声処理装置 10 は、ステップ S 4 において、セグメントのグループ化を行う。すなわち、映像音声処理装置 10 は、先のステップ S 3 において計算した非類似性測定基準と、先のステップ S 2 において抽出した特徴量とを用いて、時間的に近接して類似したセグメントを繰り返しまとめ、これらのセグメントをグループ化する。映像音声処理装置 10 は、このようにして最終的に生成されたグループを検出シーンとして出力する。

このような一連の処理を経ることによって、映像音声処理装置 10 は、ビデオデータからシーンを検出することができる。したがって、ユーザは、この結果を用いることによって、ビデオデータの内容を要約したり、ビデオデータ中の興味のあるポイントに迅速にアクセスしたりすることが可能となる。

以下、同図に示した映像音声処理装置 10 における処理を各工程毎により詳細に説明していく。

まず、ステップ S 1 におけるビデオ分割について説明する。映像音声処理装置 10 は、ビデオ分割部 11 に入力されたビデオデータを映像セグメント又は音声セグメントのいずれか、或いは可能であればその両方に分割するが、このビデオデータにおけるセグメントの境界を自動的に検出するための技術は多くのものがあり、当該映像音声処理装置 10 において、このビデオ分割方法に特別な前提要件を設けないことは上述した通りである。一方、映像音声処理装置 10 において、後の工程によるシーン検出の精度は、本質的に、基礎となるビデオ分割の精度に依存する。なお、映像音声処理装置 1

0におけるシーン検出は、或る程度ビデオ分割時のエラーを許容することができる。特に、映像音声処理装置10において、ビデオ分割は、セグメント検出が不十分である場合よりも、セグメント検出を過度に行う場合の方が好ましい。映像音声処理装置10は、類似したセグメントの検出が過度である結果である限り、一般に、シーン検出の際に検出過度であるセグメントを同一シーンとしてまとめることができる。

つぎに、ステップS2における特徴量抽出について説明する。特徴量とは、セグメントの特徴を表すとともに、異なるセグメント間の類似性を測定するためのデータを供給するセグメントの属性である。映像音声処理装置10は、映像特徴量抽出部13や音声特徴量抽出部14により各セグメントの特徴量を計算し、セグメントの特徴を表す。映像音声処理装置10は、いかなる特徴量の具体的詳細にも依存するものではないが、当該映像音声処理装置10において用いて効果的であると考えられる特徴量としては、例えば以下に示す映像特徴量、音声特徴量、映像音声共通特徴量のようなものがある。映像音声処理装置10において適用可能となるこれらの特徴量の必要条件は、非類似性の測定が可能であることである。また、映像音声処理装置10は、効率化のために、特徴量抽出と上述したビデオ分割とを同時に行うことがある。以下に説明する特徴量は、このような処理を可能にするものである。

特徴量としては、まず映像に関するものが挙げられる。以下では、これを映像特徴量と称することにする。映像セグメントは、連続する映像フレームにより構成されるため、映像セグメントから適切な映像フレームを抽出することによって、その映像セグメントの描写

内容を、抽出した映像フレームで特徴付けることが可能である。すなわち、映像セグメントの類似性は、適切に抽出された映像フレームの類似性で代替可能である。このことから、映像特徴量は、映像音声処理装置 10 で用いることができる重要な特徴量の 1 つである。この場合の映像特徴量は、単独では静的な情報しか表せないが、映像音声処理装置 10 は、後述するような方法を適用することによって、この映像特徴量に基づく映像セグメントの動的な特徴を抽出することもできる。

映像特徴量として既知のものは多数存在するが、シーン検出のためには以下に示す色特徴量（ヒストグラム）及び映像相関が、計算コストと精度との良好な兼ね合いを与えることを見出したことから、映像音声処理装置 10 は、映像特徴量として、これらの色特徴量及び映像相関を用いることとする。

映像音声処理装置 10 において、映像における色は、2 つの映像が類似しているかを判断する際の重要な材料となる。カラーヒストグラムを用いて映像の類似性を判断することは、例えば “G. Ahanger and T.D.C. Little, A survey of technologies for parsing and indexing digital video, J. of Visual Communication and Image Representation 7:28-4, 1996” に記載されているように、よく知られている。ここで、カラーヒストグラムとは、例えば HSV や RGB 等の 3 次元色空間を n 個の領域に分割し、映像における画素の、各領域での出現頻度の相対的割合を計算したものである。そして、得られた情報からは、 n 次元ベクトルが与えられる。圧縮されたビデオデータに関しても、例えば U.S. Patent #5,708,767 号公報に記載されているように、カラーヒストグラムを、圧縮データから

直接抽出することができる。

映像音声処理装置 10 では、セグメントを構成する映像におけるもともとの YUV 色空間を、色チャンネル当たり 2 ビットでサンプルして構成した、 $2^{2 \cdot 3} = 64$ 次元のヒストグラムベクトルを用いている。

このようなヒストグラムは、映像の全体的な色調を表すが、これには時間情報が含まれていない。そこで、映像音声処理装置 10 においては、もう 1 つの映像特徴量として、映像相関を計算する。映像音声処理装置 10 におけるシーン検出において、複数の類似セグメントが互いに交差した構造は、それがまとまった 1 つのシーン構造であることを示す有力な指標となる。例えば会話場面において、カメラの位置は、2 人の話し手の間を交互に移動するが、カメラは通常、同一の話し手を再度撮影するときには、ほぼ同じ位置に戻る。このような場合における構造を検出するためには、グレイスケールの縮小映像に基づく相関がセグメントの類似性の良好な指標となることを見出したことから、映像音声処理装置 10 では、元の映像を $M \times N$ の大きさのグレイスケール映像へ間引き縮小し、これを用いて映像相関を計算する。ここで、 M と N は、両方とも小さな値で十分であり、例えば 8×8 である。すなわち、これらの縮小グレイスケール映像は、 MN 次元の特徴量ベクトルとして解釈される。

さらに上述した映像特徴量とは異なる特徴量としては、音声に関するものが挙げられる。以下では、この特徴量を音声特徴量と称することにする。音声特徴量とは、音声セグメントの内容を表すことができる特徴量であり、映像音声処理装置 10 は、この音声特徴量として、周波数解析、ピッチ、レベル等を用いることができる。こ

これらの音声特徴量は、種々の文献により知られているものである。

まず、映像音声処理装置 10 は、フーリエ変換等の周波数解析を行うことによって、単一の音声フレームにおける周波数情報の分布を決定することができる。映像音声処理装置 10 は、例えば、1 つの音声セグメントにわたる周波数情報の分布を表すために、FFT (Fast Fourier Transform; 高速フーリエ変換) 成分、周波数ヒストグラム、パワースペクトル、その他の特徴量を用いることができる。

また、映像音声処理装置 10 は、平均ピッチや最大ピッチ等のピッチや、平均ラウドネスや最大ラウドネス等の音声レベルもまた、音声セグメントを表す有効な音声特徴量として用いることができる。

さらに他の特徴量としては、映像音声共通特徴量が挙げられる。これは、映像特徴量でもなく音声特徴量でもないが、映像音声処理装置 10 において、シーン内のセグメントの特徴を表すのに有用な情報を与えるものである。映像音声処理装置 10 は、この映像音声共通特徴量として、セグメント長とアクティビティとを用いる。

映像音声処理装置 10 は、映像音声共通特徴量として、セグメント長を用いることができる。このセグメント長は、セグメントにおける時間長である。一般に、シーンは、そのシーンに固有のリズム特徴を有する。このリズム特徴は、シーン内のセグメント長の変化として表れる。例えば、迅速に連なった短いセグメントは、コマーシャルを表す。一方、会話シーンにおけるセグメントは、コマーシャルの場合よりも長く、また会話シーンには、相互に組み合わされたセグメントが互いに類似しているという特徴がある。映像音声処理装置 10 は、このような特徴を有するセグメント長を映像音声共

通特徴量として用いることができる。

また、映像音声処理装置 10 は、映像音声共通特徴量として、アクティビティを用いることができる。アクティビティとは、セグメントの内容がどの程度動的或いは静的であるように感じられるかを表す指標である。例えば、視覚的に動的である場合、アクティビティは、カメラが対象物に沿って迅速に移動する度合い若しくは撮影されているオブジェクトが迅速に変化する度合いを表す。

このアクティビティは、カラーヒストグラムのような特徴量のフレーム間非類似性の平均値を測定することによって、間接的に計算される。ここで、フレーム i とフレーム j との間で測定された特徴量 F に対する非類似性測定基準を $d_F(i, j)$ と定義すると、映像アクティビティ V_F は、次式 (1) のように定義される。

$$V_F = \frac{\sum_{i=b}^{f-1} d_F(i, i+1)}{f-b} \quad \dots (1)$$

式 (1) において、 b と f は、それぞれ、1 セグメントにおける最初と最後のフレームのフレーム番号である。映像音声処理装置 10 は、具体的には、例えば上述したヒストグラムを用いて、映像アクティビティ V_F を計算することができる。

ところで、上述した映像特徴量を始めとする特徴量は、基本的にはセグメントの静的情報を表すものであることは上述した通りであるが、セグメントの特徴を正確に表すためには、動的情報をも考慮する必要がある。そこで、映像音声処理装置 10 は、以下に示すよ

うな特徴量のサンプリング方法により動的情報を表す。

映像音声処理装置 10 は、例えば図 5 に示すように、1 セグメント内の異なる時点から 1 以上の静的な特徴量を抽出する。このとき、映像音声処理装置 10 は、特徴量の抽出数を、そのセグメント表現における忠実度の最大化とデータ冗長度の最小化とのバランスをとることにより決定する。例えば、セグメント内の或る 1 画像が当該セグメントのキーフレームとして指定可能な場合には、そのキーフレームから計算されたヒストグラムが、抽出すべき特徴量となる。

映像音声処理装置 10 は、後述するサンプリング方法を用いて、対象とするセグメントにおいて、特徴として抽出可能なサンプルのうち、どのサンプルを選択するかを決定する。

ところで、或るサンプルが常に所定の時点、例えばセグメント内の最後の時点において選択される場合を考える。この場合、黒フレームへ変化 (fade) していく任意の 2 つのセグメントについては、サンプルが同一の黒フレームとなるため、同一の特徴量が得られる結果になる恐れがある。すなわち、これらのセグメントの映像内容がいかなるものであれ、選択した 2 つのフレームは、極めて類似していると判断されてしまう。このような問題は、サンプルが良好な代表値でないために発生するものである。

そこで、映像音声処理装置 10 は、このように固定点で特徴量を抽出するのではなく、セグメント全体における統計的な代表値を抽出することとする。ここでは、一般的な特徴量のサンプリング方法を 2 つの場合、すなわち、(1) 特徴量を実数の n 次元ベクトルとして表すことができる場合と、(2) 非類似性測定基準しか利用できない場合とについて説明する。なお、(1) には、ヒストグラム

やパワースペクトル等、最もよく知られている映像特徴量及び音声特徴量が含まれる。

(1) においては、サンプル数は、事前に k と決められており、映像音声処理装置 10 は、“L. Kaufman and P. J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, John-Wiley and sons, 1990” に記載されてよく知られている k 平均値クラスタリング法 (k -means-clustering method) を用いて、セグメント全体についての特徴量を k 個の異なるグループに自動的に分割する。そして、映像音声処理装置 10 は、サンプル値として、 k 個の各グループから、グループの重心値 (centroid) 又はこの重心値に近いサンプルを選択する。映像音声処理装置 10 におけるこの処理の複雑度は、サンプル数に関して単に直線的に増加するにとどまる。

一方、(2) においては、映像音声処理装置 10 は、“L. Kaufman and P. J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, John-Wiley and sons, 1990” に記載されている k -メドイドアルゴリズム法 (k -medoids algorithm method) を用いて、 k 個のグループを形成する。そして、映像音声処理装置 10 は、サンプル値として、 k 個のグループ毎に、上述したグループのメドイド (medoid) を用いる。

なお、映像音声処理装置 10 においては、抽出された動的特徴を表す特徴量についての非類似性測定基準を構成する方法は、その基礎となる静的な特徴量の非類似性測定基準に基づくが、これについては後述する。

このようにして、映像音声処理装置 10 は、静的な特徴量を複数

抽出し、これらの複数の静的な特徴量を用いることによって、動的特徴を表すことができる。

以上のように、映像音声処理装置 10 は、種々の特徴量を抽出することができる。これらの各特徴量は、一般に、単一ではセグメントの特徴を表すのに不十分であることが多い。そこで、映像音声処理装置 10 は、これらの各種特徴量を組み合わせることで、互いに補完し合う特徴量の組を選択することができる。例えば、映像音声処理装置 10 は、上述したカラーヒストグラムと映像相関とを組み合わせることによって、各特徴量が有する情報よりも多くの情報を得ることができる。

つぎに、図 4 中ステップ S 3 における特徴量を用いたセグメントの類似性測定について説明する。映像音声処理装置 10 は、2 つの特徴量について、それがどの程度非類似であるかを測定する実数値を計算する関数である非類似性測定基準を用いて、特徴量類似性測定部 17 によりセグメントの類似性測定を行う。この非類似性測定基準は、その値が小さい場合は 2 つの特徴量が類似していることを示し、値が大きい場合は非類似であることを示す。ここでは、特徴量 F に関する 2 つのセグメント S_1 , S_2 の非類似性を計算する関数を非類似性測定基準 $d_F(S_1, S_2)$ と定義する。このような関数は、以下の式 (2) で与えられる関係を満足させる必要がある。

$$\begin{aligned} d_F(S_1, S_2) &= 0 && (S_1 = S_2 \text{ のとき}) \\ d_F(S_1, S_2) &\geq 0 && (\text{全ての } S_1, S_2 \text{ について}) \\ d_F(S_1, S_2) &= d_F(S_2, S_1) && (\text{全ての } S_1, S_2 \text{ について}) \end{aligned} \quad \dots (2)$$

ところで、非類似性測定基準の中には、或る特定の特徴量にのみ適用可能なものがあるが、“G. Ahanger and T.D.C. Little, A survey of technologies for parsing and indexing digital video, J. of Visual Communication and Image Representation 7:28-4, 1996”や“L. Kaufman and P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, John-Wiley and sons, 1990”に記載されているように、一般には、多くの非類似性測定基準は、 n 次元空間における点として表される特徴量についての類似性を測定するのに適用可能である。その具体例は、ユークリッド距離、内積、 L_1 距離等である。ここで、特に L_1 距離が、ヒストグラムや映像相関等の特徴量を含む種々の特徴量に対して有効に作用することから、映像音声処理装置10は、 L_1 距離を導入する。ここで、2つの n 次元ベクトルを A 、 B とした場合、 A 、 B 間の L_1 距離 $d_{L_1}(A, B)$ は、次式(3)で与えられる。

$$d_{L_1}(A, B) = \sum_{i=1}^n |A_i - B_i| \quad \dots (3)$$

ここで、下付文字 i は、 n 次元ベクトル A 、 B のそれぞれの i 番目の要素を示すものである。

また、映像音声処理装置10は、上述したように、動的特徴を表す特徴量として、セグメントにおける様々な時点での静的な特徴量を抽出する。そして、映像音声処理装置10は、抽出された2つの動的特徴量の間類似性を決定するために、その非類似性測定基準として、その基礎となる静的特徴量の間非類似性測定基準を用い

る。これらの動的特徴量の非類似性測定基準は、多くの場合、各動的特徴量から選択された最も類似した静的特徴量の対の非類似性値を用いて決定されるのが最良である。この場合、2つの抽出された動的特徴量 SF_1 、 SF_2 の間の非類似性測定基準は、次式 (4) のように定義される。

$$d_i(SF_1, SF_2) = \min_{F_1 \in SF_1, F_2 \in SF_2} d_F(F_1, F_2) \quad \dots (4)$$

上式 (4) における関数 $d_F(F_1, F_2)$ は、その基礎となる静的特徴量 F についての非類似性測定基準を示す。なお、場合によっては、特徴量の非類似性の最小値をとる代わりに、最大値又は平均値をとってもよい。

ところで、映像音声処理装置 10 は、セグメントの類似性を決定する上で、単一の特徴量だけでは不十分であり、同一セグメントに関する多数の特徴量からの情報を組み合わせることを必要とする場合も多い。この 1 つの方法として、映像音声処理装置 10 は、種々の特徴量に基づく非類似性を、それぞれの特徴量の重み付き組み合わせとして計算する。すなわち、映像音声処理装置 10 は、 k 個の特徴量 F_1, F_2, \dots, F_k が存在する場合、次式 (5) に表される組み合わせた特徴量に関する非類似性測定基準 $d_F(S_1, S_2)$ を用いる。

$$d_F(S_1, S_2) = \sum_{i=1}^k w_i d_{F_i}(S_1, S_2) \quad \dots (5)$$

ここで、 $\{w_i\}$ は、 $\sum_i w_i = 1$ となる重み係数である。

以上のように、映像音声処理装置 10 は、図 4 中ステップ S 2 において抽出された特徴量を用いて非類似性測定基準を計算し、当該セグメント間の類似性を測定することができる。

つぎに、図 4 中ステップ S 4 におけるセグメントのグループ化について説明する。映像音声処理装置 10 は、非類似性測定基準と抽出した特徴量とを用いて、時間的に近接して類似したセグメントを繰り返しまとめてグループ化し、最終的に生成されたグループを検出シーンとして出力する。

映像音声処理装置 10 は、セグメントをグループ化してシーンを検出する際に、2 つの基本的な処理を行う。映像音声処理装置 10 は、まず第 1 の処理として、互いに時間的に近接して類似したセグメントのグループを検出する。この処理により得られるグループは、ほとんどが同一シーンの一部となるものである。そして、映像音声処理装置 10 は、第 2 の処理として、互いに時間が重複したセグメントのグループを 1 つにまとめる。映像音声処理装置 10 は、このような処理を各セグメントが独立した状態から開始し、反復して繰り返す。そして、映像音声処理装置 10 は、徐々にセグメントのグループを大きく構築していき、最終的に生成したグループをシーンの組として出力する。

このような処理において、映像音声処理装置 10 は、その処理動作を制御するために 2 つの制約を用いる。

すなわち、映像音声処理装置 10 は、第 1 の制約として、2 つのセグメントがどの程度類似している場合に、同一のシーンのものであるとみなすかを決定する非類似性閾値 δ_{sim} を用いる。例えば、図

6に示すように、映像音声処理装置10は、或るセグメントに対して、一方のセグメントが類似性領域に属するか非類似性領域に属するかを判断する。

なお、映像音声処理装置10は、非類似性閾値 δ_{sim} をユーザにより設定するようにしてもよく、また、後述するように、自動的に決定してもよい。

また、映像音声処理装置10は、第2の制約として、2つのセグメントがなお同一シーン内のセグメントとみなし得る時間軸上の隔たりの最大値として、時間閾値Tを用いる。例えば、図7に示すように、映像音声処理装置10は、時間閾値Tの範囲内で互いに近接して続いている類似した2つのセグメントA、Bを同一シーンにまとめるが、時間的に大きく離れていて時間閾値Tの範囲外である2つのセグメントB、Cをまとめることはない。このように、映像音声処理装置10は、この時間閾値Tによる時間制約があるために、互いに類似しているものの時間軸上で大きく隔たっているセグメントを同一シーンにまとめてしまうというエラーを発生することがない。

なお、この時間閾値Tとしては、6～8ショット分に相当する時間を設定した場合が概して良好な結果を与えることを見出したことから、映像音声処理装置10は、基本的に、時間閾値Tを6～8ショット単位として用いることとする。

映像音声処理装置10は、類似セグメントのグループを求めるために、ここでは、“L. Kaufman and P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, John-Wiley and sons, 1990”に記載されている階層的クラスタ分割方法 (hier

archical clustering method) を適合させて用いることにする。このアルゴリズムにおいては、2つのクラスタ C_1 、 C_2 間の非類似性測定基準 $d_c(C_1, C_2)$ について、次式(6)に示すように、それぞれのクラスタに含まれる要素間の最小非類似性として定義する。

$$d_c(C_1, C_2) = \min_{s_1 \in C_1, s_2 \in C_2} dist_s(s_1, s_2) \quad \dots (6)$$

なお、映像音声処理装置10においては、上式(6)で示される最小関数を最大関数又は平均関数に容易に置換することができる。

まず、映像音声処理装置10は、図8に示すように、ステップS11において、変数Nを初期状態のセグメントの数に初期化する。この変数Nは、常に現在検出されているグループの数を示すものである。

続いて、映像音声処理装置10は、ステップS12において、クラスタの組を生成する。映像音声処理装置10は、初期状態では、N個の各セグメントを異なるクラスタとみなす。すなわち、初期状態では、N個のクラスタが存在することになる。各クラスタは、 C^{start} と C^{end} とにより表されるその開始時と終了時とを示す特徴を有する。クラスタに含まれる要素は、 C^{start} により順序付けられたリストとして管理される。

続いて、映像音声処理装置10は、ステップS13において、変数tを1に初期化し、ステップS14において、変数tが時間閾値Tよりも大きいかな否かを判別する。ここで、映像音声処理装置10は、変数tが時間閾値Tよりも大きい場合には、ステップS23へ

と処理を移行し、変数 t が時間閾値 T よりも小さい場合には、ステップ $S15$ へと処理を移行する。ただし、ここでは、変数 t が 1 であるため、映像音声処理装置 10 は、ステップ $S15$ へと処理を移行する。

映像音声処理装置 10 は、ステップ $S15$ において、非類似性測定基準 d_c を計算し、 N 個のクラスタの中から最も類似した 2 つのクラスタを検出する。ただし、ここでは、変数 t が 1 であるため、映像音声処理装置 10 は、隣接したクラスタ間の非類似性測定基準 d_c を計算し、その中から最も類似したクラスタの対を検出する。

このような最も類似した 2 つのクラスタを検出する方法としては、対象となる全てのクラスタの対を求めることが考えられる。しかしながら、ここでは、対象とするクラスタの時間的隔たりを表す変数 t がセグメント単位で与えられ、さらにクラスタが時間順に整頓されていることから、映像音声処理装置 10 は、或るクラスタにおいて、その前後 t 個までのクラスタを非類似性の計算対象とすればよい。

ここで、検出された 2 つのクラスタをそれぞれ C_i 、 C_j と定義し、これらのクラスタ C_i 、 C_j の間の非類似性の値を d_{ij} と定義する。

映像音声処理装置 10 は、ステップ $S16$ において、非類似性値 d_{ij} が非類似性閾値 δ_{sim} よりも大きいかなんかを判別する。ここで、映像音声処理装置 10 は、非類似性値 d_{ij} が非類似性閾値 δ_{sim} よりも大きい場合には、ステップ $S21$ へと処理を移行し、非類似性値 d_{ij} が非類似性閾値 δ_{sim} よりも小さい場合には、ステップ $S17$ へと処理を移行する。ここでは、非類似性値 d_{ij} が非類似性閾値 δ_{sim} よりも小さいものとする。

映像音声処理装置 10 は、ステップ S 17 において、クラスタ C_i をクラスタ C_j に結合する。すなわち、映像音声処理装置 10 は、クラスタ C_j の要素の全てをクラスタ C_i に加える。

続いて、映像音声処理装置 10 は、ステップ S 18 において、クラスタ C_j をクラスタの組から除去する。なお、2つのクラスタ C_i 、 C_j を結合することにより開始時 C_i の値が変化した場合には、映像音声処理装置 10 は、クラスタの組の要素を開始時 C_i に基づいて再び並べ替える。

続いて、映像音声処理装置 10 は、ステップ S 19 において、変数 N から 1 を減じる。

そして、映像音声処理装置 10 は、ステップ S 20 において、変数 N が 1 であるか否かを判別する。ここで、映像音声処理装置 10 は、変数 N が 1 である場合には、ステップ S 23 へと処理を移行し、変数 N が 1 でない場合には、ステップ S 15 へと処理を移行する。ここでは、変数 N が 1 でないものとする。

すると、映像音声処理装置 10 は、ステップ S 15 において、再び非類似性測定基準 d_c を計算し、 $N-1$ 個のクラスタの中から最も類似した 2つのクラスタを検出する。ここでも、変数 t が 1 であるため、映像音声処理装置 10 は、隣接したクラスタ間の非類似性測定基準 d_c を計算し、その中から最も類似したクラスタの対を検出する。

続いて、映像音声処理装置 10 は、ステップ S 16 において、非類似性値 d_{ij} が非類似性閾値 δ_{sim} よりも大きいかな否かを判別する。ここでも、非類似性値 d_{ij} が非類似性閾値 δ_{sim} よりも小さいものとする。

そして、映像音声処理装置 10 は、ステップ S 17 乃至ステップ S 20 の処理を行う。

映像音声処理装置 10 は、このような処理を繰り返し、変数 N が減算されていった結果、ステップ S 20 において、変数 N が 1 であると判別した場合には、ステップ S 23 において、単一のセグメントのみを含むクラスタを結合する。最終的には、映像音声処理装置 10 は、この場合は、全てのセグメントが 1 つのクラスタにまとめられる形となり、一連の処理を終了する。

さて、映像音声処理装置 10 は、ステップ S 16 において、非類似性値 d_{ij} が非類似性閾値 δ_{sim} よりも大きいと判別した場合には、ステップ S 21 へと処理を移行するが、この場合には、ステップ S 21 において、時間的に重複しているクラスタを繰り返し結合する。すなわち、 C_i の時間間隔 $[C_i^{start}, C_i^{end}]$ が、 C_j の時間間隔 $[C_j^{start}, C_j^{end}]$ と相交している場合には、2 つのクラスタ C_i と C_j は、時間軸上で重複している。これにより、映像音声処理装置 10 は、クラスタをその組の開始時 C_i^{start} に基づいて整頓することによって、重複しているクラスタを検出し、それらのクラスタを 1 つに結合することができる。

そして、映像音声処理装置 10 は、ステップ S 22 において、変数 t に 1 を加算して $t = 2$ とし、ステップ S 14 へと処理を移行して変数 t が時間閾値 T よりも大きいかな否かを判別する。ここでも、変数 t が時間閾値 T よりも小さいものとし、映像音声処理装置 10 は、ステップ S 15 へと処理を移行するものとする。

映像音声処理装置 10 は、ステップ S 15 において、非類似性測定基準 d_c を計算し、現在存在する複数のクラスタの中から最も類似

した2つのクラスタを検出する。ただし、ここでは、変数 t が2であるため、映像音声処理装置10は、隣接クラスタ及び1つおきに隔たっているクラスタ間の非類似性測定基準 d_c を計算し、その中から最も類似したクラスタの対を検出する。

そして、映像音声処理装置10は、ステップS16において、隣接クラスタ及び1つおきに隔たっているクラスタ C_i 、 C_j の非類似性値 d_{ij} が非類似性閾値 δ_{sim} よりも大きいかな否かを判別する。ここでも、非類似性値 d_{ij} が非類似性閾値 δ_{sim} よりも大きいものとし、映像音声処理装置10は、ステップS21及びステップS22の処理を経て、変数 t に1を加算して $t = 3$ として再びステップS14以降の処理へと移行する。ここで、映像音声処理装置10は、変数 t が3のときには、ステップS15において、2つおきに隔たっているクラスタまでに存在するクラスタとの間の非類似性測定基準 d_c を計算し、その中から最も類似したクラスタの対を検出する。

映像音声処理装置10は、このような処理を繰り返し、変数 t が加算されていった結果、ステップS14において、変数 t が時間閾値 T よりも大きいと判別すると、ステップS23へと処理を移行し、単一のセグメントのみを含むクラスタを結合する。すなわち、映像音声処理装置10は、孤立しているクラスタを単一のセグメントのみを含むクラスタとみなし、このような一連のクラスタが存在している場合には、これらのクラスタをまとめて結合していく。この工程は、近接のシーンとは類似性関連を有さないセグメントをまとめるものである。なお、映像音声処理装置10は、必ずしもこの工程を行う必要はない。

このような一連の処理によって、映像音声処理装置10は、複数

のクラスタをまとめていき、検出シーンを生成することができる。

なお、映像音声処理装置 10 は、非類似性閾値 δ_{sim} をユーザにより設定するようにしてもよく、自動的に決定してもよいことは上述した通りである。ただし、非類似性閾値 δ_{sim} として固定値を用いる場合には、その最適値は、ビデオデータの内容に依存することとなる。例えば、変化に富んだ映像内容を有するビデオデータの場合、非類似性閾値 δ_{sim} は、高い値に設定される必要がある。一方、変化が少ない映像内容を有するビデオデータの場合、非類似性閾値 δ_{sim} は、低い値に設定される必要がある。ここで一般に、非類似性閾値 δ_{sim} が高い場合には、検出されるシーンは少なくなり、非類似性閾値 δ_{sim} が低い場合には、検出されるシーンが多くなるという性質がある。

これより、映像音声処理装置 10 においては、最適な非類似性閾値 δ_{sim} を決定することが、その性能を左右する上で重要となる。そのため、映像音声処理装置 10 においては、非類似性閾値 δ_{sim} をユーザにより設定する場合には、上述したことを考慮した上で設定する必要がある。一方、映像音声処理装置 10 は、以下に示す方法により、有効な非類似性閾値 δ_{sim} を自動的に決定することもできる。

例えば、その 1 つの方法として、映像音声処理装置 10 は、

(n) (n-1) / 2 個のセグメント対の間の非類似性の分布における平均値やメジアン（中央値）といった統計量を用いて、非類似性閾値 δ_{sim} を得ることができる。いま、全てのセグメント対における非類似性の平均値とその標準偏差をそれぞれ μ , σ とした場合、非類似性閾値 δ_{sim} は、 $a\mu + b\sigma$ の形式で表すことができる。ここで、 a 及び b は定数であり、それぞれ、0.5 及び 0.1 に設定す

ることが良好な結果を与えることを見出している。

実用上においては、映像音声処理装置 10 は、全てのセグメント対について、それらの間の非類似性を求める必要はなく、その平均値 μ 及び標準偏差 σ が真値に十分近い結果を与えるに足りるセグメント対を、全セグメント対集合からランダムに選択し、その非類似性を求めればよい。映像音声処理装置 10 は、このようにして得られた平均値 μ 及び標準偏差 σ を用いることによって、適切な非類似性閾値 δ_{sim} を自動的に決定することができる。

また、映像音声処理装置 10 は、シーンを検出する際に、セグメントが同一グループに属するか否かを決定するために、単一の非類似性測定基準を用いるばかりではなく、重み付け関数を用いて、異種の特徴量に関する多様な非類似性測定基準を組み合わせることができることは上述した通りである。映像音声処理装置 10 において、このような特徴量の重み付けは、試行錯誤の末得られるものであるが、各特徴量が質的に異なるタイプのものである場合には、通常、適切な重み付けを行うことは困難である。しかしながら、例えば、カラーヒストグラムとテクスチャフィーチャとを組み合わせる場合には、映像音声処理装置 10 は、各特徴量に関してそれぞれシーンを検出し、検出された各シーン構造を単一のシーン構造に合成することによって、両者の特徴を考慮したシーン検出を実現することができる。ここで、各特徴量に関してシーンを検出したそれぞれの結果をシーン層と称することにする。例えば、特徴量としてカラーヒストグラムとセグメント長とを用いる場合、映像音声処理装置 10 は、それぞれの特徴量に基づいたシーン検出によって、カラーヒストグラムについてのシーン層と、セグメント長についてのシーン層

とを得ることができる。そして、映像音声処理装置 10 は、これらのシーン層を単一のシーン構造に組み合わせることができる。

さらに、一般には、映像領域と音声領域とからの情報を組み合わせることはできないが、映像音声処理装置 10 は、質的に異なるタイプの特徴量に基づいた構造を組み合わせる場合と同様な方法により、映像領域と音声領域とからの情報に基づいて得られるシーン層を単一のシーン構造に組み合わせることができる。

このような処理のアルゴリズムについて説明する。ここでは、それぞれが類似性の 1 つの基準を表す k 個の特徴量 F_1, F_2, \dots, F_k があるものとし、各特徴量 F_i に対応して、非類似性測定基準 d_{F_i} と、非類似性閾値 δ_{sim} と、時間閾値 T_i とがあるものとする。映像音声処理装置 10 は、これらの各特徴量 F_i に対する非類似性測定基準 d_{F_i} と、非類似性閾値 δ_{sim} と、時間閾値 T_i とを用いて、シーン層の組 $X_i = \{X_{i,j}\}$ を検出する。例えば、映像音声処理装置 10 は、映像情報と音声情報とに対して分割的にシーン層を検出し、映像情報と音声情報とに関する 2 つの独立したシーン層 $X_i = \{X_{i,j}\}$ ($i = 1, 2$) を生成する。

映像音声処理装置 10 においては、異なるシーン層を単一のシーン構造に組み合わせるため、シーン境界の組み合わせ方を決定する必要がある。このシーン境界は、互いにそろっている保証はない。ここで、各シーン層に関して、シーン境界を示す一連の時間で表される境界点を $t_{i,1}, t_{i,2}, \dots, t_{i,|X_i|}$ で与えることとする。まず、映像音声処理装置 10 は、種々のシーン層を単一のグループに組み合わせるために、最初に或るシーン層を境界点の整列に関する基礎とするために選択する。そして、映像音声処理装置 10 は、

他のシーン層の境界が最終的に組み合わせて生成するシーン構造におけるシーン境界かどうかを各境界点 $t_{i1}, t_{i2}, \dots, t_{i|X_i|}$ に対して決定していく。

ここで、 $B_i(t)$ を、 i 番目のシーン層 X_i において、或る時間 t で近接にそのシーン層の境界点があるかどうかを示す論理関数とする。この「近接」の意味は、シーン層 X_i の状況に応じて変化し、例えば、映像情報と音声情報とのそれぞれに基づくシーン層を結合する場合には、0.5 秒程度が適当である。

映像音声処理装置 10 は、各境界点 $t_j = t_{ij}, j = 1, \dots, |X_i|$ に関して、 $l = 1, \dots, k$ のそれぞれについて、関数 $B_l(t_j)$ の結果を計算する。この結果は、それぞれのシーン層に関して、時間 t_j の近くに境界点があるかどうかを示している。そして、映像音声処理装置 10 は、結合シーン構造において時間 t_j がシーン境界であるかどうかを決定する際に、決定関数として、 $B_i(t_j)$ の値を用いる。

このような決定関数の 1 つの単純な例は、 $B_i(t_j)$ が真である個数を計数し、この個数が定数 m 以上となった場合に、結合シーン構造のシーン境界とみなすことである。特に、 $m = 1$ の場合には、全てのシーン層の境界点を最終シーン構造の境界点とみなすことと同義であり、一方、 $m = k$ の場合には、全てのシーン層において共通してシーン境界とみなされた場合のみ、結合シーン構造の境界点とすることと同義である。

このようにして、映像音声処理装置 10 は、異なるシーン層を単一のシーン構造に結合することができる。

以上説明してきたように、本発明の実施の形態として示す映像音

声処理装置 10 は、シーン構造を抽出するものである。この映像音声処理装置 10 における本手法が、例えばテレビドラマや映画等の様々な内容のビデオデータに対して、そのシーン構造を抽出可能であることは、既に実験により検証済みである。

また、映像音声処理装置 10 は、完全に自動的であり、上述した非類似性閾値や時間閾値を設定するために、ユーザの介入を必要とせず、ビデオデータの内容の変化に応じて、適切な閾値を自動的に決定することができる。

さらに、映像音声処理装置 10 は、ユーザが事前にビデオデータの意味的な構造を知る必要はないものである。

さらにまた、映像音声処理装置 10 は、非常に単純であり計算上の負荷も少ないため、セットトップボックスやデジタルビデオレコーダ、ホームサーバ等の家庭用電子機器にも適用することができる。

また、映像音声処理装置 10 は、シーンを検出した結果、ビデオブラウジングのための新たな高レベルアクセスの基礎を与えることができる。そのため、映像音声処理装置 10 は、セグメントではなくシーンといった高レベルのビデオ構造を用いてビデオデータの内容を視覚化することにより、内容に基づいたビデオデータへの容易なアクセスを可能とする。例えば、映像音声処理装置 10 は、シーンを表示することにより、ユーザは、番組の要旨をすばやく知ることができ、興味のある部分を迅速に見つけることができる。

さらに、映像音声処理装置 10 は、シーン検出の結果、ビデオデータの概要又は要約を自動的に作成するための基盤が得られる。一般に、一貫した要約を作成するには、ビデオデータからのランダム

な断片を組み合わせるのではなく、ビデオデータを、再構成可能な意味を持つ成分に分解することができることを必要とする。映像音声処理装置10により検出されたシーンは、そのような要約を作成するための基盤となる。

なお、本発明は、上述した実施の形態に限定されるものではなく、例えば、セグメント間の類似性測定のために用いる特徴量等は、上述したもの以外でもよいことは勿論であり、その他、本発明の趣旨を逸脱しない範囲で適宜変更が可能であることはいうまでもない。

産業上の利用可能性

以上詳細に説明したように、本発明にかかる信号処理方法は、供給された信号の内容の意味構造を反映するパターンを検出して解析する信号処理方法であって、信号を構成する連続したフレームのひと続きから形成されるセグメントから、その特徴を表す少なくとも1つ以上の特徴量を抽出する特徴量抽出工程と、特徴量を用いて、特徴量のそれぞれ毎に、セグメントの対の間の類似性を測定する測定基準を算出して、この測定基準によりセグメントの対の間の類似性を測定する類似性測定工程と、特徴量と測定基準とを用いて、セグメントのうち、互いの時間的距離が所定の時間閾値以内であるとともに、互いの非類似性が所定の非類似性閾値以下である2つのセグメントを検出し、信号の内容の意味構造を反映し時間的に連続するセグメントからなるシーンにまとめるグループ化工程とを備える。

したがって、本発明にかかる信号処理方法は、信号において類似したセグメントを検出してシーンにまとめることができ、セグメン

トよりも高いレベルの構造を抽出することができる。

また、本発明にかかる映像音声処理装置は、供給されたビデオ信号の内容の意味構造を反映する映像及び／又は音声のパターンを検出して解析する映像音声処理装置であって、ビデオ信号を構成する連続した映像及び／又は音声フレームのひと続きから形成される映像及び／又は音声セグメントから、その特徴を表す少なくとも1つ以上の特徴量を抽出する特徴量抽出手段と、特徴量を用いて、特徴量のそれぞれ毎に、映像及び／又は音声セグメントの対の間の類似性を測定する測定基準を算出して、この測定基準により映像及び／又は音声セグメントの対の間の類似性を測定する類似性測定手段と、特徴量と測定基準とを用いて、映像及び／又は音声セグメントのうち、互いの時間的距離が所定の時間閾値以内であるとともに、互いの非類似性が所定の非類似性閾値以下である2つの映像及び／又は音声セグメントを検出し、ビデオ信号の内容の意味構造を反映し時間的に連続する映像及び／又は音声セグメントからなるシーンにまとめるグループ化手段とを備える。

したがって、本発明にかかる映像音声処理装置は、ビデオ信号において類似した映像及び／又は音声セグメントを検出してまとめ、シーンとして出力することが可能であり、映像及び／又は音声セグメントよりも高いレベルのビデオ構造を抽出することが可能となる。

請求の範囲

1. 供給された信号の内容の意味構造を反映するパターンを検出して解析する信号処理方法であって、

上記信号を構成する連続したフレームのひと続きから形成されるセグメントから、その特徴を表す少なくとも1つ以上の特徴量を抽出する特徴量抽出工程と、

上記特徴量を用いて、上記特徴量のそれぞれ毎に、上記セグメントの対の間の類似性を測定する測定基準を算出して、この測定基準により上記セグメントの対の間の類似性を測定する類似性測定工程と、

上記特徴量と上記測定基準とを用いて、上記セグメントのうち、互いの時間的距離が所定の時間閾値以内であるとともに、互いの非類似性が所定の非類似性閾値以下である2つのセグメントを検出し、上記信号の内容の意味構造を反映し時間的に連続するセグメントからなるシーンにまとめるグループ化工程とを備えること

を特徴とする信号処理方法。

2. 上記信号とは、ビデオデータにおける映像信号と音声信号との少なくとも1つであること

を特徴とする請求の範囲第1項記載の信号処理方法。

3. 上記特徴量抽出工程によって、単一のセグメント内の異なる時点における複数の上記特徴量の単一の統計的な代表値を選択して抽出すること

を特徴とする請求の範囲第1項記載の信号処理方法。

4. 上記非類似性閾値は、複数個のセグメントの対の間の類似性の

統計値を用いて決定されること

を特徴とする請求の範囲第 1 項記載の信号処理方法。

5. 上記セグメントのうち、上記グループ化工程にてシーンにまとめられなかった少なくとも 1 つ以上のセグメントを、単一のシーンとしてまとめること

を特徴とする請求の範囲第 1 項記載の信号処理方法。

6. 上記グループ化工程により得られた任意の特徴量に関するシーン検出結果と、上記グループ化工程により得られた上記任意の特徴量とは異なる特徴量に関する少なくとも 1 つ以上のシーン検出結果とを単一にまとめること

を特徴とする請求の範囲第 1 項記載の信号処理方法。

7. 上記グループ化工程により得られた上記映像信号における少なくとも 1 つ以上のシーン検出結果と、上記グループ化工程により得られた上記音声信号における少なくとも 1 つ以上のシーン検出結果とを単一にまとめること

を特徴とする請求の範囲第 2 項記載の信号処理方法。

8. 供給されたビデオ信号の内容の意味構造を反映する映像及び／又は音声のパターンを検出して解析する映像音声処理装置であって、

上記ビデオ信号を構成する連続した映像及び／又は音声フレームのひと続きから形成される映像及び／又は音声セグメントから、その特徴を表す少なくとも 1 つ以上の特徴量を抽出する特徴量抽出手段と、

上記特徴量を用いて、上記特徴量のそれぞれ毎に、上記映像及び／又は音声セグメントの対の間の類似性を測定する測定基準を算出して、この測定基準により上記映像及び／又は音声セグメントの対

の間の類似性を測定する類似性測定手段と、

上記特徴量と上記測定基準とを用いて、上記映像及び／又は音声セグメントのうち、互いの時間的距離が所定の時間閾値以内であるとともに、互いの非類似性が所定の非類似性閾値以下である2つの映像及び／又は音声セグメントを検出し、上記ビデオ信号の内容の意味構造を反映し時間的に連続する映像及び／又は音声セグメントからなるシーンにまとめるグループ化手段とを備えること

を特徴とする映像音声処理装置。

9. 上記特徴量抽出手段は、単一の映像及び／又は音声セグメント内の異なる時点における複数の上記特徴量の単一の統計的な代表値を選択して抽出すること

を特徴とする請求の範囲第8項記載の映像音声処理装置。

10. 上記非類似性閾値は、複数個の映像及び／又は音声セグメントの対の間の類似性の統計値を用いて決定されること

を特徴とする請求の範囲第8項記載の映像音声処理装置。

11. 上記映像及び／又は音声セグメントのうち、上記グループ化手段によりシーンにまとめられなかった少なくとも1つ以上の映像及び／又は音声セグメントを、単一のシーンとしてまとめること

を特徴とする請求の範囲第8項記載の映像音声処理装置。

12. 上記グループ化手段により得られた任意の特徴量に関するシーン検出結果と、上記グループ化手段により得られた上記任意の特徴量とは異なる特徴量に関する少なくとも1つ以上のシーン検出結果とを単一にまとめること

を特徴とする請求の範囲第8項記載の映像音声処理装置。

13. 上記グループ化工程により得られた上記ビデオ信号の映像信

号における少なくとも 1 つ以上のシーン検出結果と、上記グループ化工程により得られた上記ビデオ信号の音声信号における少なくとも 1 つ以上のシーン検出結果とを単一にまとめることを特徴とする請求の範囲第 8 項記載の映像音声処理装置。

1/6

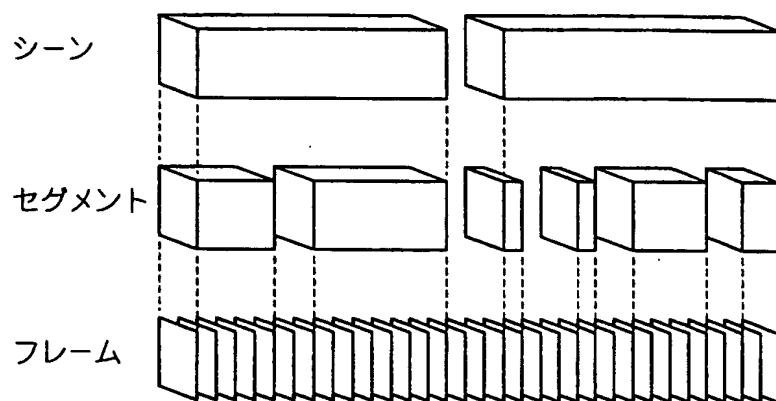


FIG. 1

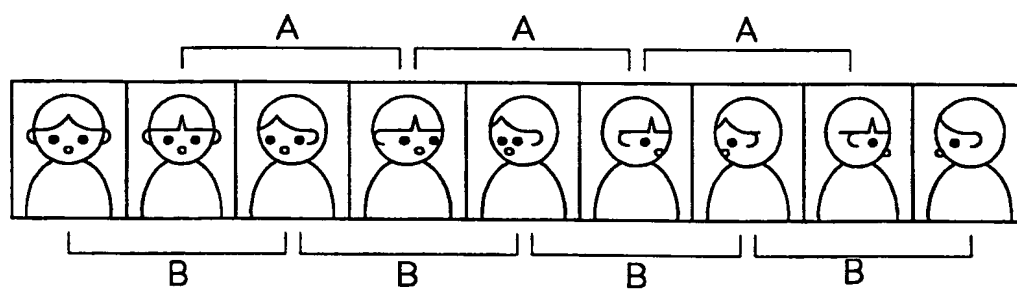


FIG. 2

THIS PAGE BLANK (USPTO)

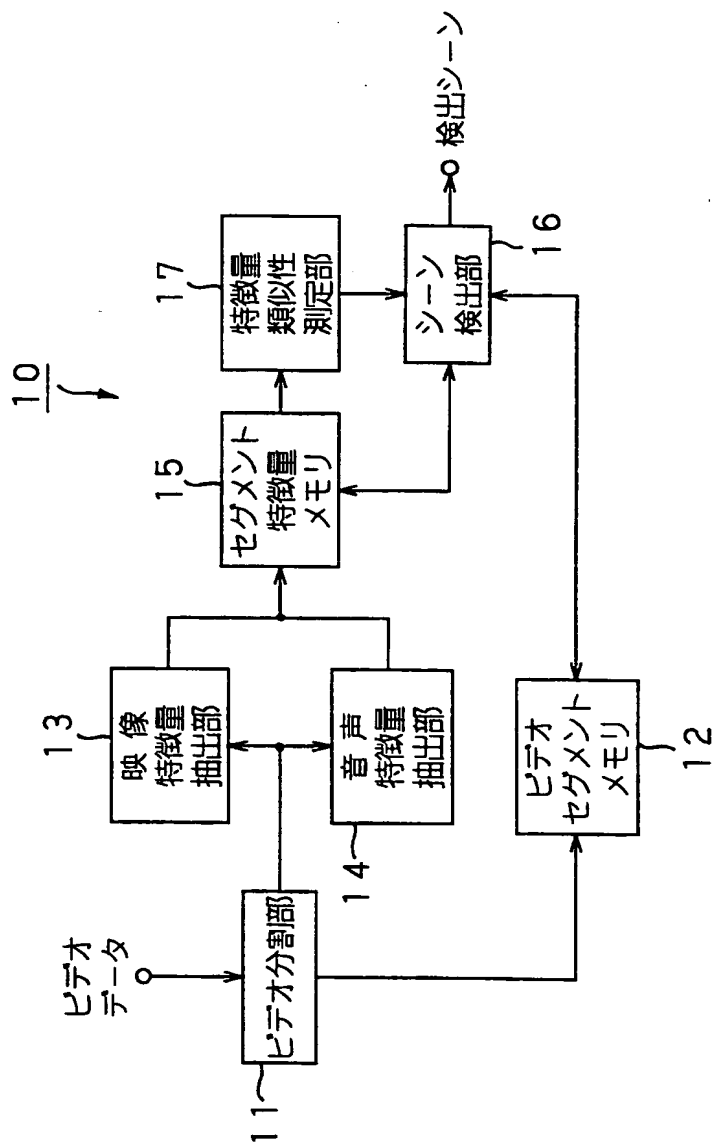


FIG. 3

THIS PAGE BLANK (USPTO)

3/6

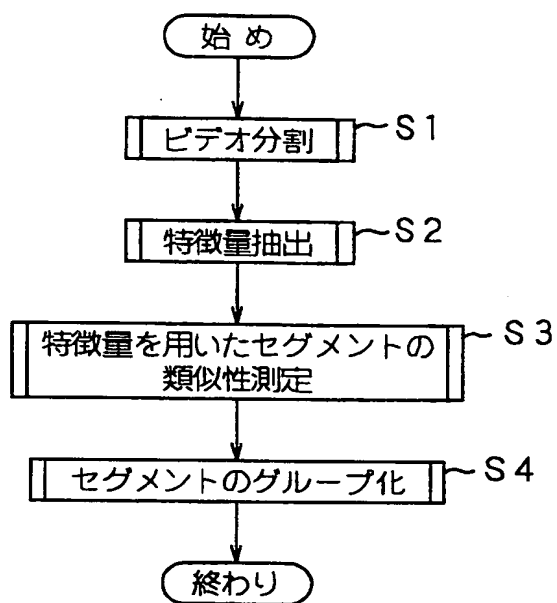


FIG. 4

THIS PAGE BLANK (USPTO)

4/6

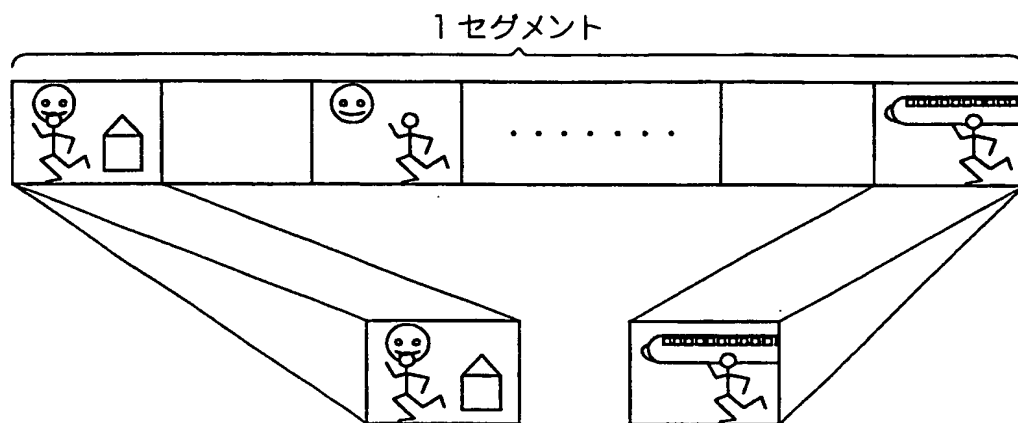


FIG. 5

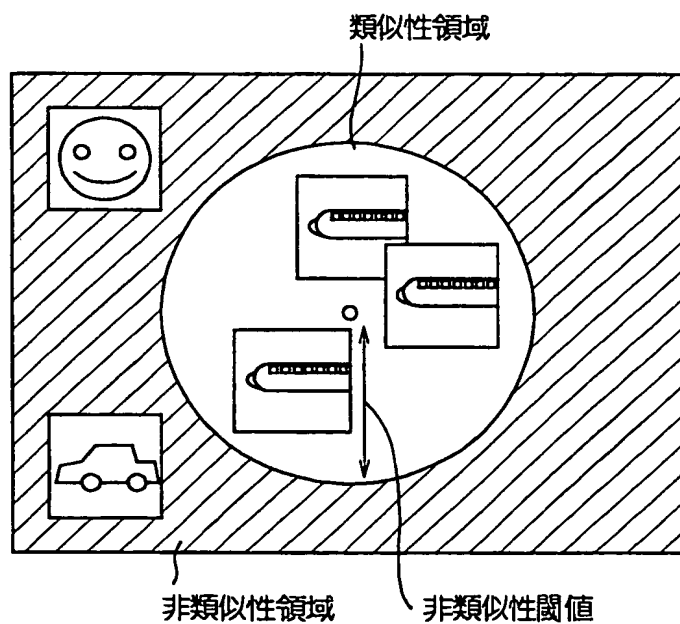


FIG. 6

THIS PAGE BLANK (USPTO)

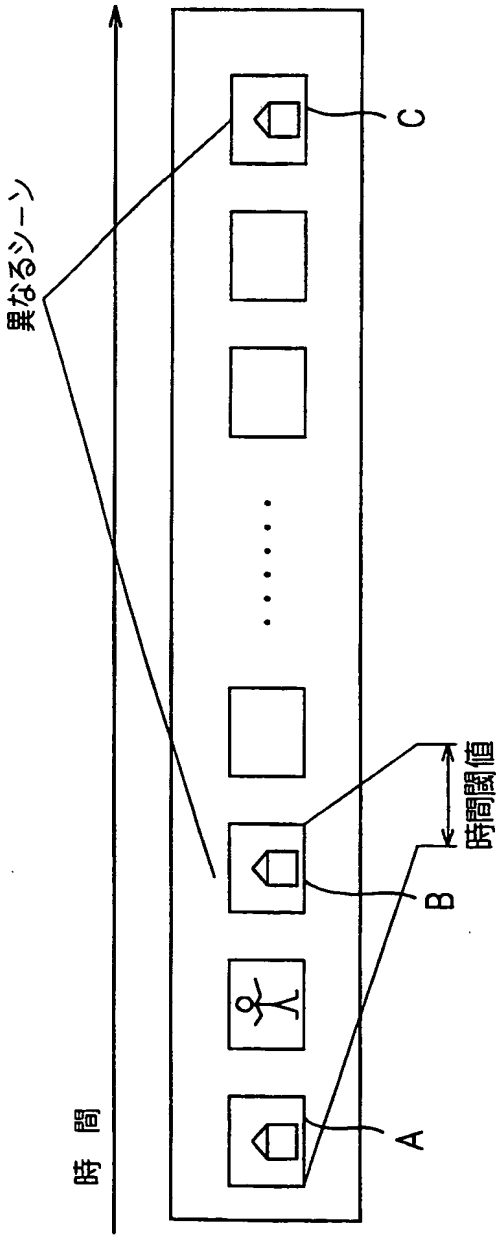


FIG. 7

THIS PAGE BLANK (USPTO)

6/6

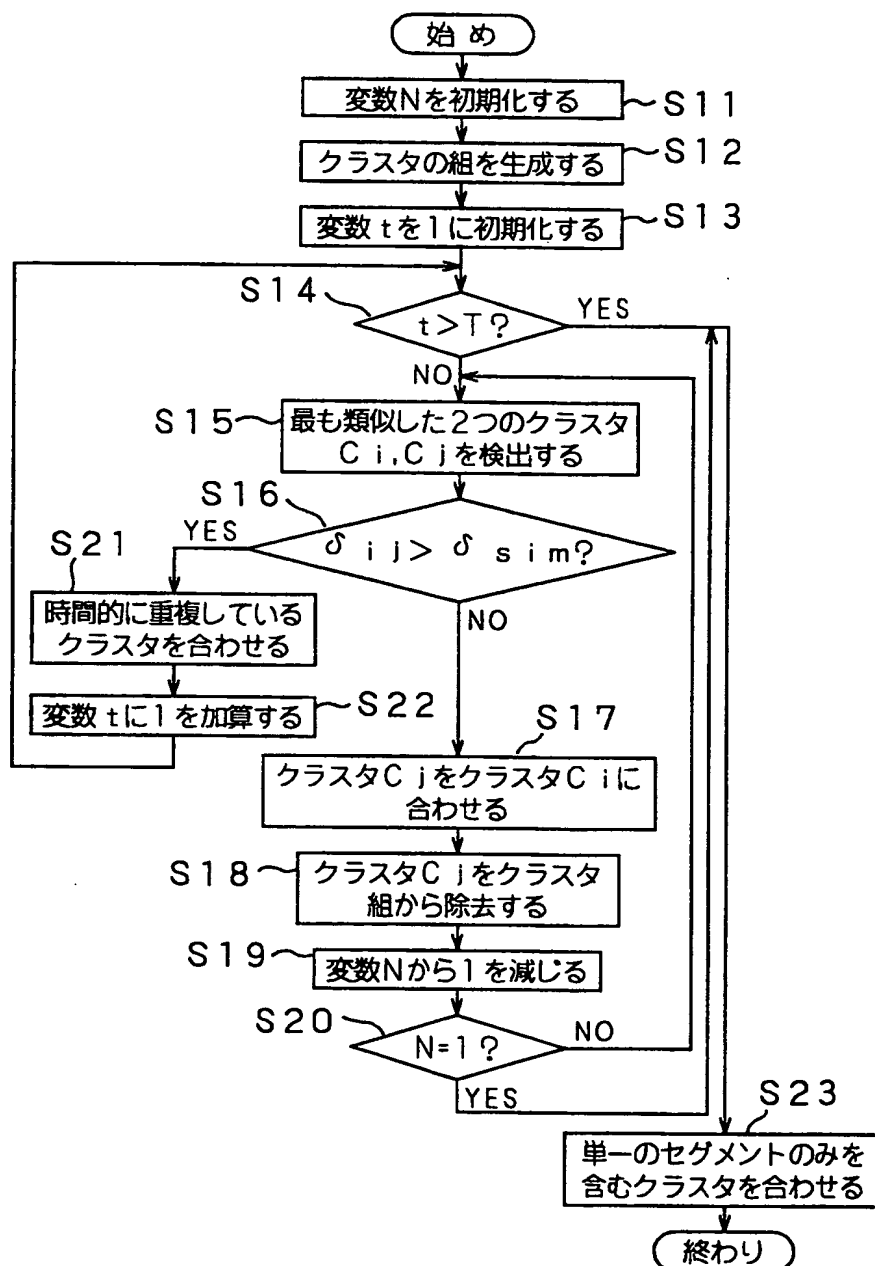


FIG. 8

THIS PAGE BLANK (USPTO)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP00/00423

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl.⁷ H04N7/32
 Int.Cl.⁷ H04N5/92
 Int.Cl.⁷ G06T7/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl.⁷ H04N5/91~5/956 Int.Cl.⁷ H04N7/24~7/68
 Int.Cl.⁷ H04N5/222~5/257 Int.Cl.⁷ G06T7/00
 Int.Cl.⁷ H04N5/262~5/28

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
 Jitsuyo Shinan Koho 1922-1996 Toroku Jitsuyo Shinan Koho 1994-2000
 Kokai Jitsuyo Shinan Koho 1971-2000 Jitsuyo Shinan Toroku Koho 1996-2000

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP, 10-257436, A (Atsushi Matsushita, Kenichi Okada), 25 September, 1998 (25.09.98), Full text; Figs. 1 to 20 (Family: none)	1-13
A	JP, 7-193748, A (Nippon Telegr. & Teleph. Corp. <NTT>), 28 July, 1995 (28.07.95), Full text; Figs. 1 to 14 (Family: none)	1-13
A	EP, 711078, A (Matsushita Electric Ind. Co., Ltd.), 08 May, 1996 (08.05.96), Full text; all drawings & JP, 8-181995, A & US, 5751377, A	1-13

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search
 18 April, 2000 (18.04.00)

Date of mailing of the international search report
 02 May, 2000 (02.05.00)

Name and mailing address of the ISA/
 Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

THIS PAGE BLANK (USPTO)

国際調査報告

国際出願番号 PCT/JPO0/00423

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl⁷ H04N7/32Int. Cl⁷ H04N5/92Int. Cl⁷ G06T7/00

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl⁷ H04N5/91~5/956Int. Cl⁷ H04N7/24~7/68Int. Cl⁷ H04N5/222~5/257Int. Cl⁷ G06T7/00Int. Cl⁷ H04N5/262~5/28

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1922-1996年

日本国公開実用新案公報 1971-2000年

日本国登録実用新案公報 1994-2000年

日本国実用新案登録公報 1996-2000年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	JP, 10-257436, A (松下温, 岡田謙一) 25. 9月. 1998 (25. 09. 98) 全文, 第1-20図 (ファミリーなし)	1-13
A	JP, 7-193748, A (日本電信電話株式会社) 28. 7月. 1995 (28. 07. 95) 全文, 第1-14図 (ファミリーなし)	1-13

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの

「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)

「O」口頭による開示、使用、展示等に言及する文献

「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&」同一パテントファミリー文献

国際調査を完了した日

18. 04. 00

国際調査報告の発送日

02.05.00

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

安田 太

5H

2955

電話番号 03-3581-1101 内線 3531

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	EP, 711078, A (松下電器産業株式会社) 8. 5月. 1996 (08. 05. 96) 全文, 全図 & JP, 8-181995, A & US, 5751377, A	1-13